ELECTROLYTIC GROWTH PROCESSES WITH APPLICATIONS TO ADAPTIVE SYSTEMS

By

W. A. AINSWORTH.

Thesis presented for the degree of Ph.D.

in the University of Keele.

Department of Communication,
University of Keele,
Staffordshire.

September, 1963.

ABSTRACT

Two types of electrolytic growth processes have been investigated as possible means of controlling impedances in adaptive systems. One of these entails the growth of conductors by metallic deposition and the other the growth of insulation by anodic oxidation.

The way in which the growth of a metallic dendrite changes the impedance of a cell has been determined, and it has been found that this depends very much on the shape of the cell. By restricting growth to narrow channels it has been found possible to reliably control the impedance of cells. The properties of devices employing this principle are described.

The growth of insulating films on aluminium has been found to be another satisfactory method of reproducibly adjusting the impedance of a cell. The properties of these films and the impedance changes which take place during their 'formation' and 'erosion' are discussed. These films have special properties which make them suitable for use in the fabrication of devices containing many variable impedances in a single unit. The construction and properties of these devices are described.

An adaptive system employing the growth of dendrites has been constructed, and trained to distinguish between four simple patterns. The design of this machine and the training procedure employed are given.

Suggestions are made for future work and for other applications of electrolytic growth processes.

ACKNOWLEDGMENTS

I should like to express my gratitude to Professor D.M. MacKay for suggesting the research topic and for many helpful discussions during the course of the work; and to Miss B.G. Brammer for typing this thesis.

- - - - - - - -

CONTENTS

- - - - - - - - - - - -

CHAPTER I

1. (1) Introductory

Since the advent of the electronic computer there has been a great deal of speculation about the similarity between large calculating machines and human brains. The differences are numerous and obvious, but this comparison has had the effect of leading engineers and biologists along a common, stimulating line of research. They have attempted to build machines which simulate the various activities of animal and human brains. This has been done partly in order to understand better these activities and the brains themselves, and partly in an effort to replace human operators in arduous and boring tasks. Out of these studies has emerged a new scientific discipline which has come to be known as "Cybernetics". This has been defined by Wiener (1948) as the study of control and communication in the animal and the machine.

1. (2) Cybernetics.

Each new scientific discipline evolves new concepts. In order to refer to these concepts existing words with less specialised meanings are often used. The same word is sometimes used by different persons to refer to similar but different concepts with the result that a certain amount of confusion is generated. Such a situation,

which is probably inevitable in the formative stages of a new discipline, has occurred in some sections of the literature of cybernetics. For this reason the sense in which some of the special terms are used in this thesis will be described in the next few sections.

1 (3) Adaptive systems. *

One of the most striking features of the human brain, and other biological systems, is its ability to modify its behaviour to cope with the situation in which it finds itself. There appear to be two types of events which give rise to this ability. Firstly, there are the temporary changes brought about automatically by the changes in environmental conditions. For example, if the air temperature rises the pores of the skin open so that more moisture evaporates and the body temperature remains constant. Secondly, there are the permanent changes which occur when learning takes place.

Systems which can modify their internal state automatically in order to keep some parameter constant, or perhaps optimal, will be called self-adaptive systems. An example of such a system is the controller of a manufacturing process. The controller has little control over certain variables, such as the cost of raw materials,

*The term 'system' is very general. In this thesis it is used to refer to anything being discussed as though it were a machine. This usage limits the term as it implies that a system has a purpose.

but it can adjust others - temperatures and flow rates - in order to keep production at a maximum, or costs at a minimum.

1 (4) Self-organising systems

The permanent changes in the behaviour of systems are called learning only if these changes lead to an improvement in the ability of the system to carry out its purpose. A system which adjusts its internal state in such a way as to bring about learning will be called a self-organising system.(Ashby, 1947).

Several attempts have been made to define a self-organising system but none of these are really adequate. (Yovits and Cameron, 1960; v Foerster and Zopf, 1962; Yovits et al., 1962) Some are general but so vague that it is difficult to tell from them exactly what constitutes a self-organising system. Others are more precise but in such a way that they seem to attribute to self-organising systems properties which they would not otherwise have possessed. This being the case, it seems best for the present to keep in mind a few examples which are generally accepted as being self-organising systems, and to decide whether other systems fall into this category by analogy with these examples. This alternative is more acceptable than using an unsatisfactory definition which may lead to logical difficulties.

1 (5) Evolutionary systems

The term self-organising system was first introduced with reference to the functioning of the brain, so brains are looked upon as the prime examples of self-organising systems. There is, however, another fundamental example which serves better as an introduction because the mechanism by which it operates is much more clearly understood. This is the process of evolution (Pringle, 1951; Pask, 1961; Bremermann, 1962) The goal of this process, or system, is the survival of the species.

The mechanism by which it acts may be summarised as follows. Some of the offspring differ from their parents in certain respects. If these changes help the animal to survive better, then they are likely to produce offspring which have these characteristics which aid survival. The changes take place at random so most of them will be useless and individuals possessing them will not survive or reproduce even as well as their parents.

Such a process is not altogether useful in a stationary environment. If the parents survive in that environment at least long enough to produce offspring, it is quite sufficient for the offspring to be exact reproductions of their parents in order that the race should survive for ever. But if the environment changes

it is necessary for the creatures which inhabit it to change too. This process - natural selection - makes it possible for this to happen.

From the above, it can be seen that all that is required of a self-organising system of the evolutionary kind is that there should be a system which has a goal. Changes take place in the system which affect its performance with respect to this goal. If these changes improve performance they are retained; if they cause deterioration they are eliminated. This occurs automatically when the goal is the natural one of survival, but it must be done artificially if other goals are to be attained.

1 (6) Artificial self-organising systems.

The general purpose digital computer can be used to simulate any system, so it has naturally been used to embody self-organising systems. A specific example of a computer being employed as a self-organising system is one which is learning to play a game. A number of computer programmes have been devised which play games such as noughts and crosses (Michie, 1961.) draughts (Samuel, 1961) and even chess (Shannon, 1950; Bronowski, 1950). These programmes include in them the rules of the game and methods of choosing the best move out of the possible moves at any stage in the game. This decision

procedure may be entirely statistical. In the simple case of noughts and crosses it is possible to store the probability of ultimate success for each move given the state of the game. This has been done with a programme in which the probabilities were adjusted at the end of each game until the optimum move was picked each time. (v. Michie, 1961).

For more complicated games it is not feasible to store probabilities for each move because machines are not likely to be large or fast enough. There are also more efficient schemes. Strategies are stored and these are modified as a result of their effectiveness in actual play.

In these artificial self-organising systems there exist the same essential features as in natural ones : a goal (to win the game or as many games as possible), a means of evaluating performance (knowing when a piece has been taken and when the game has been won or lost), and a way of altering behaviour (by modifying probabilities or programme sub-routines).

1(7) Limitations of computers.

General purpose computers are built to solve any computational problem and must therefore be inefficient at any specific task compared with a machine built specially for that task. Having seen that learning can be caused to take place in machines, the next step is to

build special machines which use the self-organising principle in order to perform their tasks.

Self-organising and self-adaptive systems require some internal parameters which can be adjusted. In digital computers only discrete changes can be made. These changes can be as small as desired if many-bit words are used to represent each parameter. For a machine of a given size, however, the number of parameters which can be represented is inversely related to the number of bits in each word, so the requirements of many parameters and small changes are antagonistic. It appears that it would be more efficient to build a machine in which the parameters could be represented by continuously-variable quantities.

There is some encouragement for supposing this in a widely-held theory of how information is stored, permanently, in the central nervous system (Bullock, 1959). Brains are certainly more efficient with respect to size at storing information than electronic computers.

In computers the calculations take place in one part whilst the information used in these calculations is stored in another part. In brains it seems probable that these two activities are diffused throughout the whole. It would seem that systems with brain-like properties, such as self-organisation, should follow the example of the nervous system and store their information as widely as

possible. Continuously-variable devices are the most suitable vehicles for this (MacKay, 1958; 1962).

1 (8) Continuously-variable devices.

There are in existence a number of continuously-variable devices, some of which have already been employed in machines which exhibit learning properties. The most widely used of these is the motor-driven rheostat which provides a continuously-variable resistance. These devices are too large and expensive to be used in machines which contain many thousands of variable parameters, but they have been employed in the "perceptron" of Rosenblatt (1960).

A simple form of continuously-variable quantity is the charge on a condenser. There is another requirement, however, which excludes condensers from a list of useful adaptive components. It is desirable that the value of a variable parameter should not change spontaneously. Even the very best condensers do not have infinite leak resistance so the voltage on the plates will be subject to an inevitable decay. Calculations show that about 12 hours is the practical limit for the time constant of condensers.

Another continuously-variable device is the "solion". This device has an output voltage which may be made proportional to the integral with respect to time of the current through it. It consists of an electrochemical cell

with two compartments. The current transfers ions from one compartment to another. This sets up a concentration potential which forms the output voltage. This device suffers from the same difficulty as the condenser. The ions tend to diffuse so as to produce a uniform concentration, and the output voltage decays.

There are a number of magnetic components which may be used as continuously-variably devices. These make use of the hysteresis loop of ferromagnetic materials. Some of these devices have been designed specifically for use in adaptive machinery (Brain, 1960; Hawkins et al, 1961). The problem with their employment is that they require expensive, auxiliary electronic circuits.

1 (9) Electrolytic growth processes.

Another possible physical realisation of such a device is an electrolytic cell whose impedance may be modified by a growth process at one of the electrodes. The impedance change may be brought about either by the growth of a conductor which reduces the effective distance between the electrodes, or by the growth of insulation between the electrodes.

Electrochemical devices have a number of properties which, it is thought, make them suitable for employment in adaptive systems. They can be made quite small and are, therefore, economical of power, and probably extremely cheaply.

Another factor which should be taken into consideration when self-organising systems are being designed is the time in which the changes in the parameters must be made. The time in which a significant change in the impedance of an electrolytic cell with electrode separation of about 1cm. may be brought about is of the order of seconds. This is reduced if the physical dimensions are too. Such values seem extremely suitable.

1 (10) Organisation of the thesis.

The object of this thesis is to give an account of the physical properties of examples of both types of electrolytic growth processes, and to explain how these processes may be incorporated into devices which may be used to determine the variable parameters of adaptive systems.

In Chapter II the factors which affect the growth of metallic, dendritic crystals are discussed, and a number of experiments which were performed to see how this growth could be controlled are described. The results of this investigation led to the construction of a number of devices. The design, construction and properties of these devices are described in Chapter III.

In the following chapter, Chapter IV, the physical properties of the electrolytic growth of anodic oxide films on aluminium are discussed.

The factors affecting the breakdown or such films and the effect this has on the impedance of the cell are also dealt with. Chapter V is analogous to Chapter III in that it describes the incorporation of the physical processes dealt with in the previous chapter into actual devices.

Chapter VI is a review of the work by previous authors on adaptive systems. Many of these systems may be described as conditional probability computers. This forms a useful link between intuitive ideas on adaptive systems and actual engineering design.

Chapter VII describes an adaptive system which has been built using the devices of Chapter III and the design principles elucidated in Chapter VI.

[redacted]

A summary of the most important parts of the work and the conclusions which may be drawn from the results are given in Chapter VIII. The directions which future work might take and some possible applications are also discussed.

References, Chapter 1

Ashby, W.R.(1947) "Principles of the Self-Organising Dynamic System."
J. Gen. Psychol., vol. 37, p. 125.

Brain, A.E. (1960) "The Simulation of Neural Elements by Electrical Circuits Based on Multi-Aperture Cores."
Research Report, Stanford Research Institute, California.

Bremermann, H.J. (1962) "Optimization through Evolution and Recombination" in "Self-Organizing Systems, 1962"
Spartan Books, Washington, pp. 93 - 106.

Bronowski, J (1950) "Chess Haying Machines."
Nature, vol. 166, p. 1040.

Bullock, T.H. (1959) "Neuron Doctrine and Electrophysiology."
Science, vol. 129, no. 3355.

von Foerster, H, and Zopf, G (Eds) (1962) "Principles of Self-Organization", Pergamon Press, New York.

Hawkins, J.K. Mansey C.J. and Stafford, R.A.(1961) "A Magnetic Integrator for the Perceptron Program".
Summary Report, Aeronutronic Research Lab. Publ. No. U - 1405. Newport Beach, California.

MacKay, D.M. (1958) "Operational Aspects of Intellect" in "Mechanization of Thought Processes". H.M.S.O., pp 37-52.

MacKay, D.M. (1962) "Self-Organization in the Time Domain" in "Self-Organizing Systems, 1962" Spartan Books, Washington, pp. 37-48.

Michie, D (1961) "Trial and Error.
Penguin Science Survey, vol. 2.

Pask, G. (1961) "The Cybernetics of Evolutionary Processes and Self-Organising Systems".
Third Namur Conference on Cybernetics.

Pringle, J.W.S. (1951) "On the Parellel between Learning and Evolution".
Behaviour, vol. 3, p. 174.

Rosenblatt, F. (1960) "Mark I Perceptron Operators' Manual" Report No. VG-1196 - G - 5, Cornell Aeronautical Laboratory, Buffalo, New York.

Samuel, A.L. (1961) "Some Studies in Machine Learning Using the Game of Checkers" in "Lernende Automaten", Oldenburg, Munich, pp 155-178.

Shannon, C.E. (1950) "Programming a Computer for Playing Chess." Phil. Mag., Sr. 7, vol. 41, pp. 256 - 275.

Wiener, N. (1948) "Cybernetics".
Wiley, New York.

Yovits, M.C. and Cameron, S. (Eds.) (1960) "Self-Organizing Systems", Pergamon Press, New York.

Yovits, M.C. et al. (Eds.) (1962) "Self-Organizing Systems, 1962",.
Spartan Books, Washington.

CHAPTER II

THE ELECTROLYTIC GROWTH OF DENDRITES

2 (1) Introductory.

The principle of the electrodeposition of metals from solutions of their salts has been suggested by a number of investigators as a process possessing features suitable for modelling biological systems and for the construction of adaptive machinery. A survey of some of their devices will be given later; this chapter is concerned with the physical and chemical properties of cells in which such processes take place.

2 (2) Electrodeposition of metals

When a metal is immersed in an electrolyte an equilibrium tends to be established in which a steady potential difference exists between the metal and the solution. The magnitude of this potential difference depends upon the concentration of the electrolyte. It is brought about by metal ions dissolving into the solution, leaving an excess of electrons in the metal. The metal electrode is then negatively charged and so attracts the positive metal ions in the solution. A dynamic equilibrium is reached when the electrode potential is such that the number of metal ions entering and leaving the metal are equal. There is a tendency for dissolved metal ions to cluster

round the surface of the electrode and this in turn attracts the excess electrons in the metal to the surface. There is thus a negatively charged region inside the metal and a positively charged one in the solution near the electrode. This region is known as the electrical double layer.

If a potential different from equilibrium potential is maintained between the electrode and the electrolyte by an external source of power, the metal ions may be dissolved or deposited continuously, subject, of course to the supply of ions being maintained. In order to complete the external circuit it is necessary to introduce a second electrode. The total arrangement is then an electrolytic cell. If this second electrode is of the same metal as the first it is possible to dissolve ions from one (the anode) and deposit them on the other (the cathode).

2 (3) Factors affecting the type of deposit.

Metal is deposited on the cathode of an electrolytic cell in the form of crystals. Each new ion arriving at the cathode can either be added to one of the already present crystals and so cause it to grow, or it can form the nucleus of a new crystal. The physical conditions determine which of these situations is more likely to come about. If the former process occurs a few large crystals will be formed, whereas the latter will produce many small crystals.

An accumulation of small crystals has the physical appearance of a smooth, uniform deposit. This is normally the result which is desired in electroplating. This process has been extensively studied and the factors which produce this type of deposit are fairly well known (e.g. v. Potter, 1956).

The conditions which favour the formation of a few, larger, crystals were probably first found accidentally when electroplaters were unsuccessful in producing smooth, uniform deposits. They have since, however, been studied in their own right, in particular by Fischer and his collaborators (Fischer and Heiling, 1954; Fischer, 1954). The factors which influence the physical nature of electro-deposited metals are named by Glasstone (1943) as current density, the nature and concentration of the electrolyte, the presence of inhibitors, the temperature of the electrolyte, and the nature of the cathode material.

The presence of an inhibitor reduces the tendency for crystals to grow and so encourages nucleation. The result is an aggregate of many small crystals spread uniformly over the cathode. Added impurities normally act as inhibiting agents, espectially if they contain colloidal matter. Other factors, however, such as the concentration of the anion and the temperature of the electrolyte may act as inhibitors. Fischer and Heiling (1954) mention that inhibition increases as the concentration

of the anion is increased, and decreases as the temperature of the electrolyte is increased.

The effects of current density and electrolyte concentration are in many ways complementary. In a cell where inhibition is slight the tendency is for the crystal to grow into regions where the ion concentration is greatest. If the electrolyte concentration is small the metal ions in the neighbourhood of the cathode will rapidly be removed, unless the current density is also small. The crystals will then grow outwards towards the richer regions. If the electrolyte concentration is raised it is possible to raise the current density to a greater level before the cathodic region becomes depleted in metal ions and the crystals grow outwards. Other factors being equal, large currents and low concentrations favour the formation of few crystals growing outwards in a direction perpendicular to the cathode, while small currents and high concentrations produce a deposit consisting of small crystals spread evenly over the cathode.

The nature of the negative ion in the solution affects the cathodic deposit in many ways, but in general it seems that simple ions favour the production of large crystals and complex ions small ones. This is presumably due to the inhibitory effect of the complex anion.

Increasing the temperature of the elctrolyte decreases

inhibition and so favours the growth of large crystals. It also, however, increases the rate of diffusion and so compensates for any ion depletion in the cathodic region due to the low electrolyte concentration and high current density. The net effect of small changes in temperature is thus negligible.

The nature of the cathode is important with respect to the adhesion of the deposit, but does not affect the physical form of the deposit once it has been covered by a few molecular layers.

2 (4) Dendritic growth

The resistance of a simple, electrolytic cell increases with the separation of the electrodes and decreases with the area of the electrodes and the conductivity of the electrolyte. One of the simplest methods of varying the resistance of a cell is to alter the effective separation of the electrodes by growth at one of them.

The metal ions being deposited during this growth process must come from the electrolyte. If the number of ions in the electrolyte is reduced its conductivity will be decreased, so the net effect of the growth process might be an increase in the resistance of the cell. This possibility may be avoided by ensuring that each ion deposited on the cathode is replaced by one dissolved from the anode. This situation can only occur if the anode and

cathode are of the same metal.

If the electrodes are of the same size, and if the anode dissolves and the cathode grows uniformly the distance between the two will remain constant. It is necessary to introduce some asymmetry into the cell, such as a large anode and a small cathode, if a predictable change in resistance is to occur.

Alternatively, the conditions in the cell can be arranged so that a non-uniform growth takes place at the cathode (Unless the anode is specially shaped it is difficult to get anything but uniform dissolution there.) The conditions which favour the growth of crystals outwards from the surface of the cathode are low inhibition, low metal ion concentration, and high current density.

The type of growth produced by these conditions resembles the branches of a tree. For this reason crystals of this form are known as "dendrites".

2 (5) Growth in open dishes.

Some experiments were performed to see how these general considerations applied to a few particular cells which had previously been suggested as variable resistance elements (MacKay, 1958). A cell was constructed of a dish containing an aqueous $^{N}/_{20}$ solution of silver nitrate in which were immersed two silver electrodes. The electrodes were of 32 S.W.G. wire and were set about 3mm. apart. A

piece of filter paper was placed between the electrodes to provide a substrate on which dendrites could be grown. The space between the electrodes was viewed with a low power microscope.

A potential of 2 volts was applied across the cell and it was observed that the cell resistance was about 60 kilohms. A dendrite grew slowly in the general direction of the anode and eventually made contact with it, reducing the resistance of the cell to about 8 ohms. Most of this resistance change took place, however, as the dendrite made contact with the anode.

A number of dendrites were grown in similar cells, and it was observed that not all of them grew towards the anode. The ones which grew approximately in this direction turned towards it as they approached it. Also the rate of growth increased as the tip of the dendrite came near to the anode. This suggested that dendrites would grow more consistently and more rapidly in a stronger electric field.

2 (6) Growth of dendrites in alcoholic solution.

If the voltage applied across the type of cell described above is raised much above 2 volts the potential across the double layer is great anough for hydrogen ions to cross it and the evolution of bubbles of hydrogen gas to occur at the cathode. This interferes with the formation of the dendrites. If some solvent other than water is

employed this limitation need not apply.

It was known that dendrites could be grown in an electrolyte consisting of a solution of stannous chloride in capryl alcohol (Addison, 1959). Preliminary experiments confirmed that this was so. The potential between the electrodes was first increased to 4 volts. No gases were evolved and spiky growths were observed to form on the cathode. These grew in straight lines and could be made to branch by interrupting the flow of current for a few seconds. The direction of growth was, once more, not always towards the anode.

Capryl alcohol has a very low conductivity. The initial resistance of the cell was about 4 megohms when it contained a solution of 25 gm. of stannous chloride per litre of capryl alcohol compared with 60 kilohms for an N/20 (8.5 gm. per litre) aqueous silver nitrate solution. This made it possible to increase the voltage across the cell by a large factor without a large current flowing or the evolution of hydrogen.

The voltage was increased to 120 volts and a dendrite, which resembled a thread, grew fairly rapidly from cathode to anode. On reaching the anode it broke into many pieces and then reformed. The dendrite acted rather like a fuse. When it had grown sufficiently to make contact with the anode a large current passed which destroyed the connection.

In order to limit the maximum current through the cell a one megohm resistor was connected in series with it. This enabled a dendrite to be grown which was not destroyed completely when it made contact with the anode. Only the tip broke off. During the formation of the dendrite the resistance of the cell dropped from about 4 megohms to about 0.5 megohms. It did not remain at this latter value but fluctuated around it.

2 (7) Effect of electrolyte concentration on crystal size.

Dendrites were grown in capryl alcohol containing different concentrations of stannous chloride. It was observed that when 60 volts were applied across a cell containing 25 gm. per litre stannous chloride, dendrites which resembled threads were formed. When the same potential was applied across the same cell but containing a solution of concentration 50 gm. per litre, straight dendrites with orthogonal spikes were grown.

The same type of difference in the physical appearance of the dendrites was observed when the concentration of the electrolyte was kept constant and the applied voltage varied. With the 25 gm. per litre solution the spike-like dendrites grew when the applied potential was 4 volts, and the thread-like dendrites grew when a potential of 120 volts was applied across the electrodes.

When these various growths were viewed at higher

magnification it was observed that the differences in the physical appearance were due to differences in the size of the crystals grown under different conditions. The thread-like dendrites were quite similar to the spike-like ones when they were magnified about ten times more.

2 (8) Effect of the strength of the electric field.

Whenever a dendrite was grown it was observed that when the tip of a dendrite came close to the anode it kept breaking and then reforming. In this situation the end of the dendrite is in a rather strong electric field. In order to determine whether this phenomenon was related to the strength of the electric field the potential across the whole cell was increased so that the dendrite was in a higher electric field throughout its formation.

A cell was constructed which initially had a field of about 500 volts per cm. between the electrodes. In a typical experiment a dendrite began to grow from the cathode rather rapidly. Suddenly the end piece broke off and drifted towards the anode. The part from which this had broken then began to grow. Sometimes it grew so rapidly that it caught up with the drifting section and joined on to it. The original tip then once more became the growing point. This cycle of events continued, the dendrite breaking and then reforming.

This phenomenon was probably caused by the electrostatic

force on the dendrite. The dendrite and the anode may be considered to be two plates of a condenser, and the low conductivity electrolyte to be the dielectric. This condenser is charged when the dendrite grows and so the dendrite will experience a force. If this force is great enough the dendrite will break and then regrow from the break point.

In the absence of an electric field it was observed that these tin dendrites dissolved in the solution. This dissolution did not take place just at the tip, as the growth did, but along the whole surface of the dendrite. If this also happens when a field is applied dendrites continually get thinner. The strength in one region of the dendrite will then, sooner or later, become insufficient to sustain the electrostatic force and it will break.

2 (9) Inadequacies of dendrites grown in open dishes.

These preliminary studies showed that although it is possible to change the resistance of an electrolytic cell by the growth of dendrites, it is not possible to do this in a predictable or repeatable way in open dishes. There are three main difficulties. The first of these is that the dendrites did not always grow directly between the electrodes. It seems that there are several factors which influence the direction of growth of a dendrite. The electric field is probably the most important

when it is greater than 100 volts per cm., but with lower fields other factors may predominate. Layton (1954) considers that the concentration of metal ions to be the prime factor which determines the direction of growth. This depends on the electrolyte concentration and the temperature as well as the electric field.

One of the desirable characteristics of a variable resistance is that it should be possible to adjust it in either direction. It was found that the resistance of a cell in which a dendrite had been grown could be returned to its initial value by reversing the polarity of the applied voltage. The resistance of the cell could not be increased gradually, however, as the dendrite dissolved not only at the tip but all along its length, leading to the formation of mechanical weak spots and rapid fragmentation. This is the second difficulty.

The third difficulty is the non-linear variation of resistance with the growth of a dendrite. Most of the resistance change took place as the growing tip approached the anode. This was due to the geometry of the cell. It is difficult to calculate analytically how the resistance between two wire electrodes immersed in a circular dish depends on their distance apart, so the form of this variation was determined by actually measuring the resistance as a function of distance in a

cell with variable electrode separation.

This last difficulty does not preclude the employment of such cells as variable resistance devices but it makes their use more limited.

2 (10) Growth in narrow tubes.

(a) Theoretical

It is possible to circumvent these three difficulties if the dendrite is grown in a narrow capillary tube. This forces the dendrite to grow straight; it gives the possibility of small, local resistance changes in either direction; and the shape of the cell is such that a linear relationship holds between the resistance of the cell and the length of the dendrite.

Consider a narrow tube, of cross-sectional area a, with a cathode inserted in one end and the other connected to a much wider tube. The anode is immersed in the electrolyte in the wide tube as shown in Figure (2.1).

The resistance R of this cell is given by:

$$R = R_o + \frac{l_o}{\sigma a} \qquad (2.1)$$

where l_o is the length of the tube, σ is the conductivity of the electrolyte, and R_o is the resistance of the electrolyte between the anode and the near end of the narrow tube.

If q coulombs of charge are passed through the cell, the mass of metal deposited on the cathode is given by

$$M = \frac{e}{F} q \qquad (2.2)$$

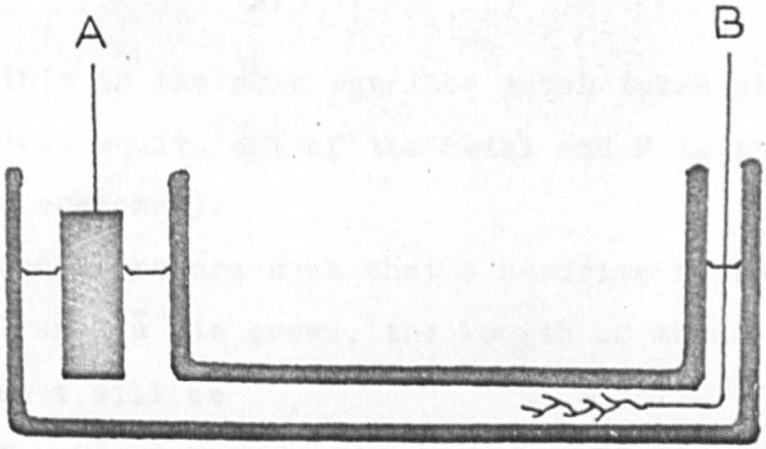


Figure 2.1.

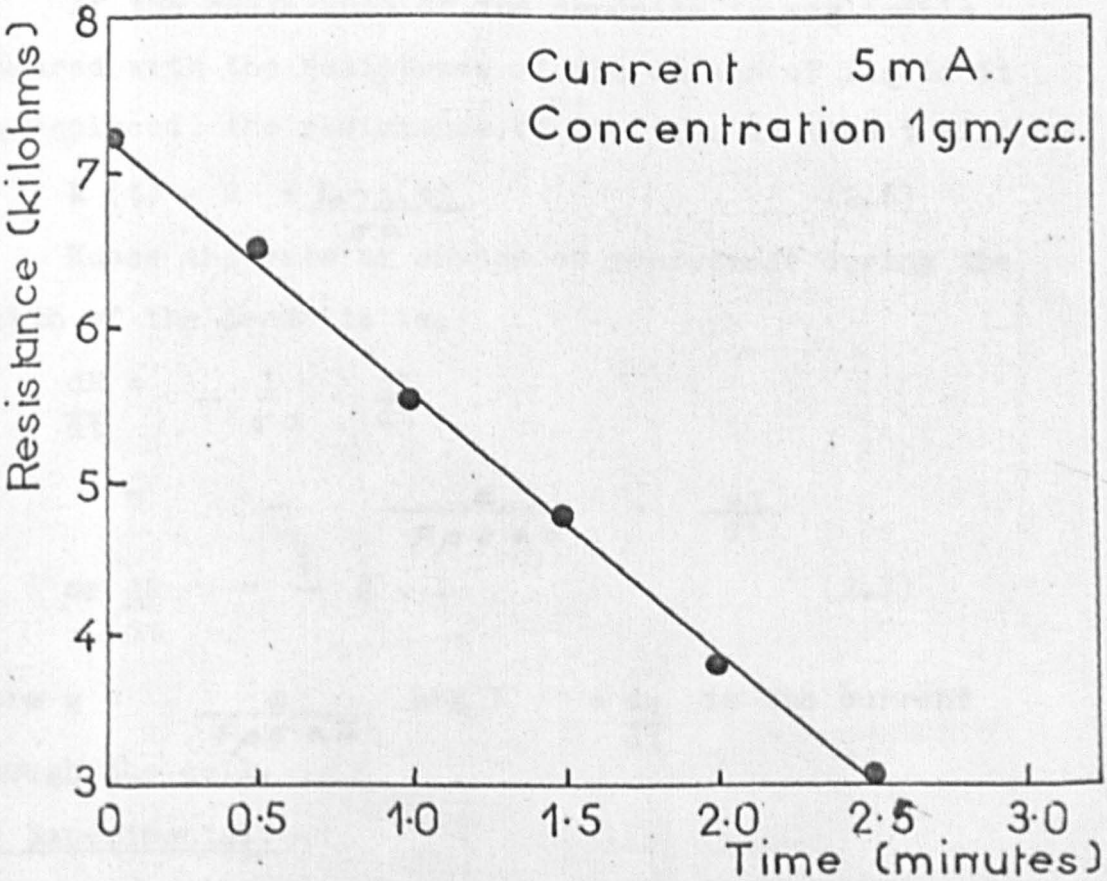


Figure 2.3.

provided that this is the only reaction which takes place. e is the chemical equivalent of the metal and F is the Faraday (96,500 coulombs).

If the conditions are such that a dendrite of mean cross-sectional area $\bar{a}$ is grown, the length of this dendrite at time t will be

$$l(t) = \frac{M}{\rho \bar{a}} \qquad (2.3)$$

where ρ is the density of the metal. This may be written,

$$l(t) = \frac{e\,q}{F \rho \bar{a}} \qquad (2.5)$$

If the resistance of the dendrite is negligible compared with the resistance of the column of liquid it has replaced, the resistance of the tube at time t will be

$$R(t) = R + \frac{l_0 - l(t)}{\sigma a} \qquad (2.6)$$

Hence the rate of change of resistance during the growth of the dendrite is,

$$\frac{dR}{dt} = -\frac{1}{\sigma a} \cdot \frac{dl}{dt}$$

$$= -\frac{e}{F \rho \sigma a \bar{a}} \cdot \frac{dq}{dt}$$

$$\text{or } \frac{dR}{dt} = -g \cdot I \qquad (2.7)$$

where $g = \frac{e}{F \rho \sigma a \bar{a}}$ and $I = \frac{dq}{dt}$ is the current through the cell.

(b) Experimental.

A cell similar to that shown in Figure (2.1) was

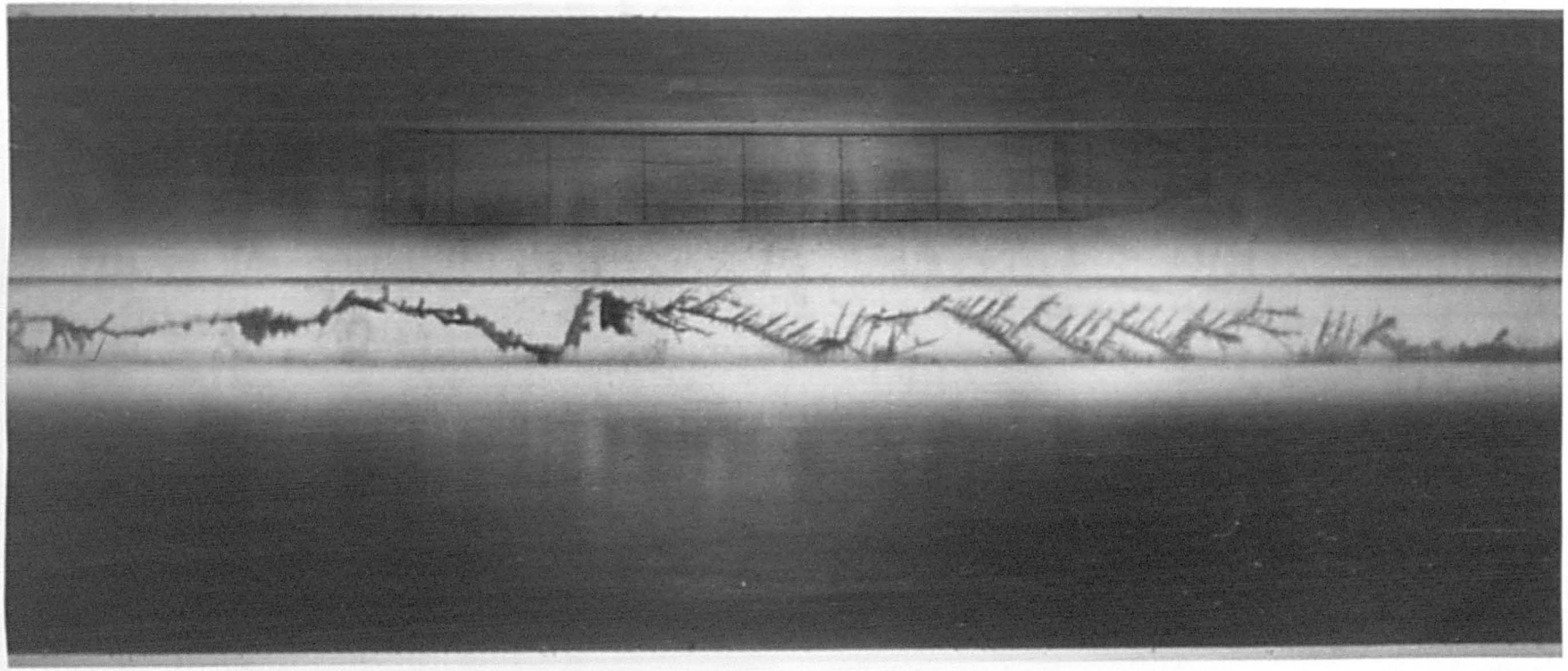


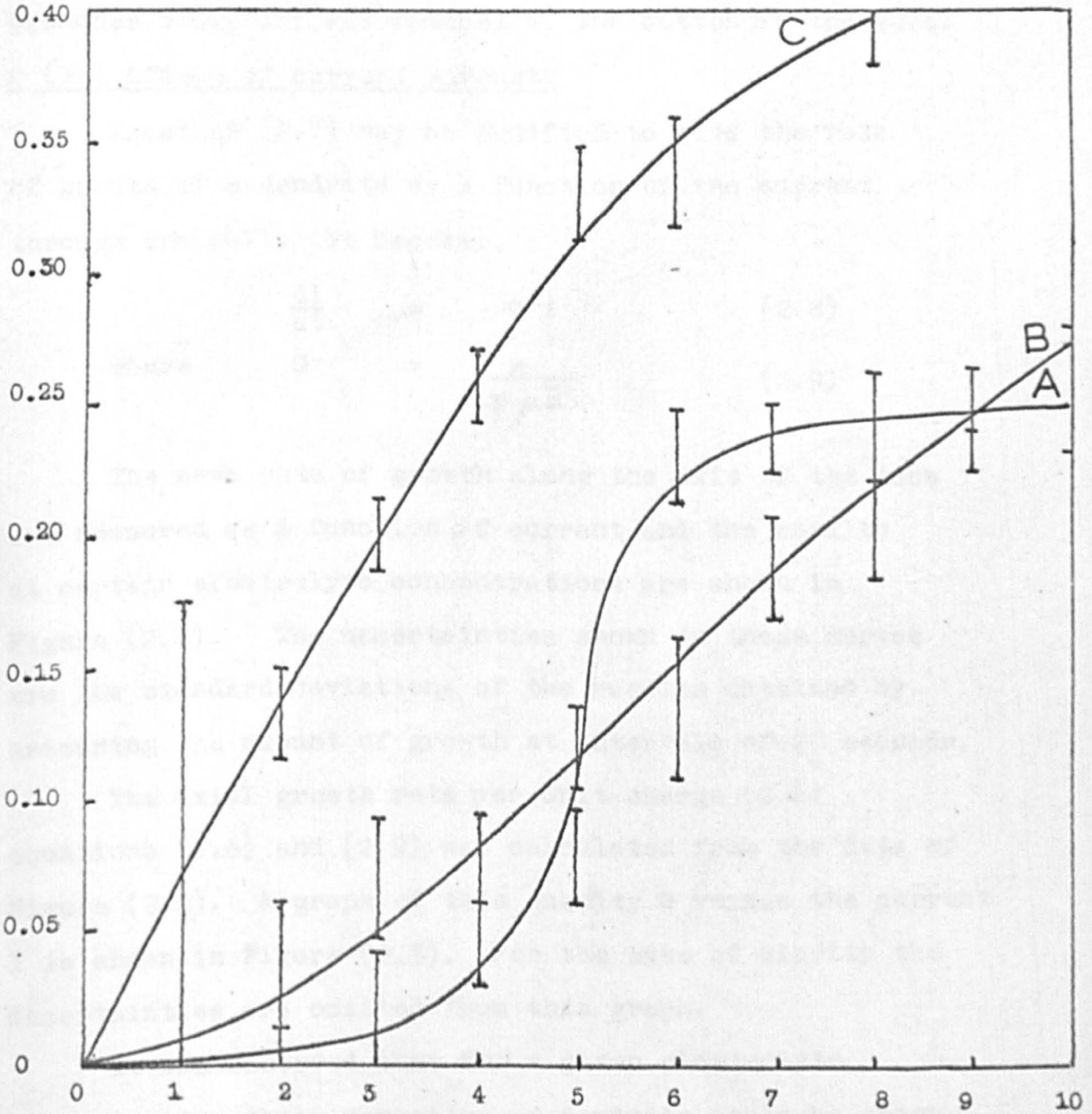
Figure 2.2.

constructed. The capillary tube was about 10 cm. long and had an internal diameter of 1 mm. The electrolyte employed was aqueous silver nitrate, and the electrodes were of silver. A current of 5 milliamps was passed through the cell and a dendrite similar to that shown in Figure (2.2) was grown.

The resistance was measuredduring the growth process. A plot of resistance versus time is shown in Figure (2.3). It was approximately a straight line of negative slope as predicted by equation (2.7). On a larger scale there was some deviation from linearity. This occurred because the dendrite did not grow regularly down the centre of the tube but meandered from side to side forming many branches.

It was found that when the current was reversed the resistance increased with time in a linear manner. This probably happened because the potential difference between the metal and solution only exceeded the dissolution potential in the region of the tip. The relatively high ohmic resistance of the electrolyte in the tube ensured that the potential of the electrolyte in contact with the rest of the dendrite was little different from equilibrium potential.

The rate of change of resistance during dissolution was slightly greater than during formation. This was due to the silver dendrite not dissolving completely. Small



RATE OF GROWTH (mm.per second) versus CURRENT (milliamps) for various electrolyte concentrations.

A 1.0 gm.per c.c. silver nitrate.
B 0.5 gm.per c.c. "
C 0.1 gm.per c.c. "

(Uncertainties correspond to a sampling time of 20 seconds.)

Figure 2.4.

branches broke off and remained at the bottom of the tube.

2 (11) Effect of current strength

Equation (2.7) may be modified to give the rate of growth of a dendrite as a function of the current through the cell. It becomes,

$$\frac{dl}{dt} = G\,i \qquad (2.8)$$

where $$G = \frac{e}{F \rho \bar{a}} \qquad (2.9)$$

The mean rate of growth along the axis of the tube was measured as a function of current and the results at certain electrolyte concentrations are shown in Figure (2.4). The uncertainties shown on these curves are the standard deviations of the results obtained by measuring the amount of growth at intervals of 20 seconds.

The axial growth rate per unit charge (G of equations (2.8) and (2.9) was calculated from the data of Figure (2.4). A graph of this quantity G versus the current I is shown in Figure (2.5). For the sake of clarity the uncertainties are omitted from this graph.

It was observed that for a given electrolyte concentration three varieties of dendrite could be grown in the tube. With reference to the 1.0 gm. per c.c. curve of Figure (2.5) these varieties corresponded approximately to the regions 0 - 4 millamps, 4-7 milliamps, and greater than 7 milliamps. In the low current region thick dendrites which almost filled the tube were grown. In the intermediate

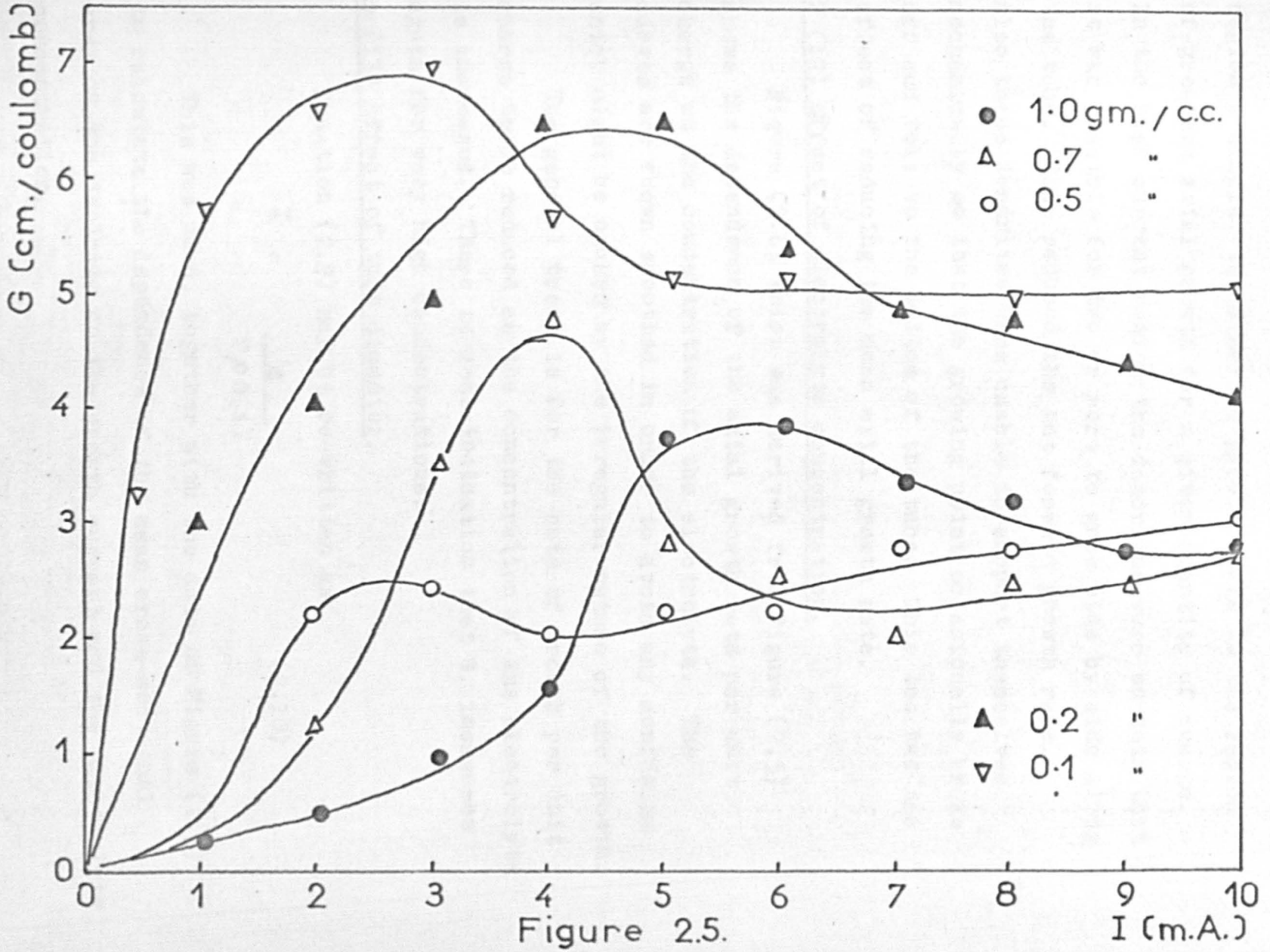


Figure 2.5.

region a single, thin dendrite grew. This is the region of greatest axial growth for a given quantity of charge. In the high current region the dendrites were so thin that it was possible for two or more to grow side by side along the tube. This reduced the net forward growth rate. Also these dendrites were unable to support themselves mechanically so that the growing point occasionally broke off and fell to the bottom of the tube. This too had the effect of reducing the mean axial growth rate.

2 (12) Effect of electrolyte concentration.

Figure (2.6) which was derived from Figure (2.5) shows the dependence of the axial growth rate per unit charge on the concentration of the electrolyte. The curves are shown smoothed in order to avoid any confusion which might be caused by the irregular nature of the growth.

The general trend is for the rate of growth per unit charge to be reduced as the concentration of the electrolyte is increased. There is some indication that G. increases again for very high concentrations.

2 (13) Effect of tube diameter.

Equation (2.9) may be re-written as:

$$\bar{a} = \frac{e}{F \rho G(i)} \qquad (2.10)$$

This was used, together with the data of Figure (2.5) to calculate the dependence of the mean cross-sectional area of the dendrite on the growth current and the electrolyte concentration.

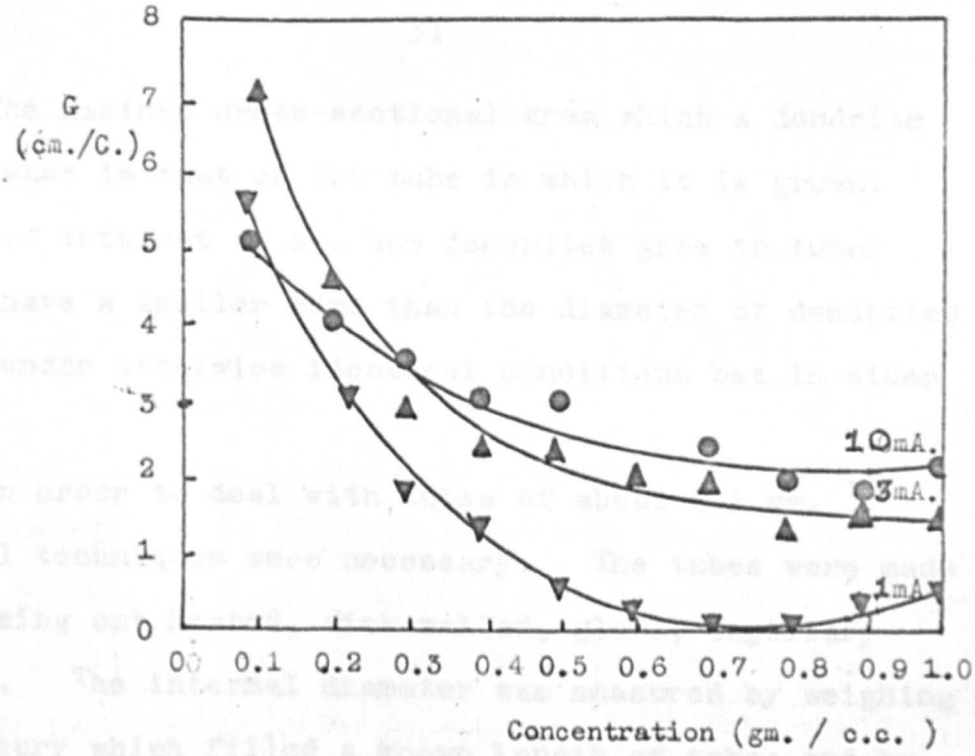


Figure 2.6.

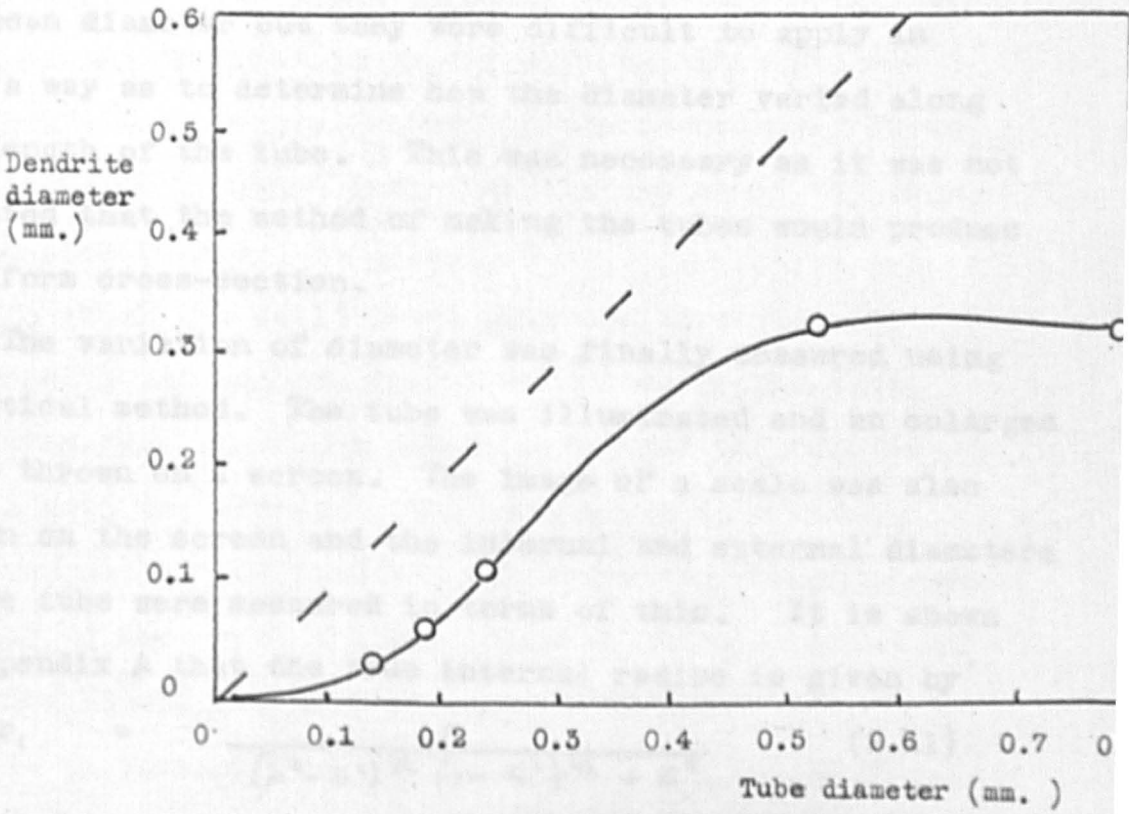


Figure 2.7.

The maximum cross-sectional area which a dendrite may assume is that of the tube in which it is grown. It is of interest to see how dendrites grow in tubes which have a smaller bore than the diameter of dendrites grown under otherwise identical conditions but in wider tubes.

In order to deal with tubes of about 0.1 mm. special techniques were necessary. The tubes were made by drawing out heated, thick-walled, glass, capillary tubing. The internal diameter was measured by weighing the mecury which filled a known length of tube; and by measuring the impedance of a length of tube filled with electrolyte of known conductivity. These methods gave the mean diameter but they were difficult to apply in such a way as to determine how the diameter varied along the length of the tube. This was necessary as it was not expected that the method of making the tubes would produce a uniform cross-section.

The variation of diameter was finally measured using an optical method. The tube was illuminated and an enlarged image thrown on a screen. The image of a scale was also thrown on the screen and the internal and external diameters of the tube were measured in terms of this. It is shown in Appendix A that the true internal radius is given by

$$r_1 = \frac{r_1'}{(\mu^2 - \bar{m}^2)^{1/2} (1 - \bar{m}^2)^{1/2} + \bar{m}^2} \qquad (2.11)$$

where r_1' = measured internal radius

$\bar{m}$ = $\frac{\text{measured internal radius}}{\text{measured external radius}}$. (2.11)

μ = refractive index of the glass.

The rate of growth was measured in a number of different tubes having diameters 0.1 to 1.0 mm. A solution of 1.0 gm. / c.c. and a current of 1 milliamps was employed. From these measurements the mean diameter of the dendrite was calculated. These results are shown in Figure (2.7). The dashed line shows the maximum diameter which a dendrite grown in a given tube could have.

2 (14) Effect of applied magnetic field.

A factor which might possibly affect the path taken by a growing dendrite is an external magnetic field. This could conceivably produce an electromagnetic force on the growth tip, and influence the motion of the ions in the electrolyte.

In tubes the walls leave only one direction available for growth, and in open dishes dendritic growth is so irregular that it is not possible to tell if magnetic fields have any effect. The advantages of a tube, however, may be retained and yet leave the dendrite some freedom of choice in the direction in which it grows if a tube containing a branch point is employed.

It is rather difficult to produce a Y - shaped tube

of narrow bore, but it is comparatively simple to construct a narrow groove of this shape in the surface of an insulating material such as perspex. This, filled with electrolyte, produces a cell which serves quite as well as a tube in constraining the growth.

A Y - shaped, three- electrode cell was constructed in this manner. The groove, about 0.5 mm. wide and deep, was filled with a saturated solution of silver nitrate. (If a more dilute solution were used evaporation rapidly caused the solution to become saturated.)

In the first series of experiments the electrodes in two of the limbs, A and B, were connected to the positive terminal of a potential source and the other electrode, C, to the negative side. This caused a dendrite to grow from C which split into two when it reached the branch point. A small hole (about 1.5 mm. diameter) was drilled at the branch point to give the dendrite some freedom of growth. If an applied magnetic field had any influence it could bias the growth from one side of this hole so that the branching dendrite entered one of the limbs, A or B, before the other. This would reduce the resistance of this limb relative to the other causing more current to pass so that the dendrite in this limb grew faster. This would be a self-accelerating process.

A magnetic field of about 2,000 oersteds was applied to the cell in a direction perpendicular to the plane of

Direction of magnetic field	Maximum growth in A	Maximum growth in B	Equal growth
+	5	3	2
−	1	7	2
None	4	5	1

Table (2.1.)

Direction of magnetic field	Maximum growth in A	Maximum growth in B	Equal growth
+	0	2	1
-	0	2	1
None	1	2	0

Table (2.2.)

the Y. A current of about 2 milliamps was passed through each limb. These currents were adjusted with an external resistance so that they were initially of the same strength. In a few preliminary runs the direction of the magnet field did seem to determine the direction in which the greatest growth took place. (Appendix D) On a later occasion, however, this was found to be only a statistical relationship. Table (2.1) shows the number of times the dendrite grew most in each limb with the field in the two directions and with no applied field.

A second series of experiments were performed in which A and B were made cathodes, and C the anode. With the other conditions similar to those in the experiment described above, dendrites grew from A and B. The results in Table (2.2) show that the magnetic field had no effect on the growth.

In view of the fact that there was some correlation between the direction of the magnetic field and the direction of growth in the first series of experiments but not in the second it appears that the electromagnetic force on the tip probably influenced the direction of growth rather than the ionic motion. However, this is only a tentative conclusion.

In order to obtain more data about the effect it is desirable to vary some of the parameters. It was

found that the behaviour of the dendrite was approximately the same in cells in which the angle between the limbs A and B was 30° and 180°. Altering the other parameters presents some difficulty. For currents less than about 2 milliamps the growth in the hole at the branch point is not dendritic so that it becomes filled, uniformly, with silver. With greater currents the electrolyte evaporates more rapidly, and the cell becomes open-circuited before the process is completed. The size of the hole cannot be changed much either. Smaller holes do not give the dendrite enough freedom so the branches grow equally in each limb, and larger holes approximate to the conditions in open dishes. It was not possible to see if the magnetic field had any effect on dendrites grown in capryl alcohol as with a saturated solution of stannous chloride the grooves had a resistance of about 500 megohms per cm. Only extremely small currents could be passed and the solution evaporated before any growth took place.

The strength of the magnetic field is a parameter which could be varied. It is thought that a probabalistic relation was found because the effect of the magnetic field was of the same order as the natural irregularities in the growth. A stronger magnetic field may produce a deterministic relationship. Such fields,

however, would be difficult to produce in miniature, low-power devices.

2.(15) Discussion of results.

Most of the experiments described in this chapter may be considered as part of an attempt to answer the question of what determines the shape to which a crystal will grow. This is a question which has been asked many times before. The first answers appeared in terms of thermodynamics. It was thought at one time that the shape to which a crystal would grow was the one which minimised the surface free energy. This approach was developed by Curie (1885) and by Wulff (1901). Later it was shown that this did not hold for crystals larger than about a micron, as above this size the departure from minimum free energy caused by non-equilibrium shape is less than that necessary to cause the crystal to grow.

Crystals grow by the addition of ions to their faces. It is known that an ion will not be bound very strongly to a face unless it is in contact with an edge or step. Because of the differences between the observed and calculated growth rates of crystals it was postulated that real crystals had a permanent supply of steps on their faces due to dislocations in the crystal. Subsequently edge and screw dislocations were observed (Frank, 1958).

The mechanism of the growth of a crystal seems to be that of an ion being brought to a crystal face, by

electrolysis or by a diffusion process, and then moving over the face until it reaches a step. The direction in which a crystal grows may thus be influenced as much by the ions within the crystal as by the external electric field or concentration gradient. It is, therefore, hardly surprising that there was a large variation in the way in which the dendrites grew under seemingly similar conditions.

A theory of the formation of electrolytic whiskers has been given by Price et al (1958). Electrolytic whiskers are formed by the electrodeposition of metals but, unlike dendrites, have a uniform diameter. Price et al consider that the metal ions arrive at the growth face by electric transport and impurities arrive there by diffusion. They assume that growth cannot take place if the impurity concentration exceeds a certain critical value. This is shown to set a limit on the radii of the whiskers. The calculated results they obtain are in very good agreement with the experimental findings.

The impurities in this instance act as inhibitors of crystal growth. In the experiments described in this chapter no impurities were deliberately introduced into the electrolyte. The anions, however, may act as inhibitors (Fischer, 1954). If this is so the degree of inhibition should increase as the concentration is increased. This may explain why the initial gradients

of the curves in Figure (2.5) are inversely related to the concentration of the electrolyte. As the current is increased the silver ions in the cathode region becomes in excess of the nitrate ions so that their inhibitory action becomes less and other factors determine the shape of the dendrite.

REFERENCES - CHAPTER II

Addison, A. (1959) Research Report, University of Illinois.

Curie, J. (1885) Bull. Soc. min. de France, vol. 8, p. 145.

Fischer, H. and Heiling H.F. (1954) "Morphology of the Growth of Isolated Crystals in Cathodic Metal Deposits." Trans. Inst. Met. Finishing, vol. 31, pp. 90-102.

Fischer, H. (1954) "Electrolytishe Abscheidung und Electrokrystallization von Metallen", Springer.

Frank, F.C. (1958) "Growth and Perfection of Crystals". Doremus, Roberts and Turnbull, Eds.), John Wiley, pp 3-9.

Glasstone, S. (1942) "Introduction to Electrochemistry," von Nostrand, Chapter 14.

Layton, D.N. (1954) Discussion in Trans. Inst. Met. Finishing, vol 31, pp. 103-105.

MacKay, D.M. (1958) Discussion in "Mechanisation Thought Processes" H.M.S.O. pp. 924-5.

Potter, E.C. 1956) "Electrochemistry," Cleaver-Hume Press, Chapter 13.

Price, P.B., Vermilyea, D.A., and Webb, M.B. (1958) "On the Growth and Properties of Electrolytic Whiskers." Acta Metallurgica, vol 6, no. 8.

Wulff, G. (1901), Z. Krist., vol. 34, p. 449.

CHAPTER III

ELECTROCHEMICAL, VARIABLE-RESISTANCE DEVICES

3. (1) Historical

During the last decade there have been a number of attempts to employ electrochemical reactions in adaptive devices. One of the earliest of these was by Drechsler (1951), who built a machine which simulated the Pavlovian reflexes found in animals. He used a tank containing an electrolyte of lead acetate and cellulose triacetate dissolved in acetic acid. The 'conditioning' of the machine caused electrical impulses to flow between electrodes immersed in this tank. These in turn caused metallic threads to grow between the electrodes, altering the conduction paths and, hence the 'behaviour' of the machine.

In 1957, MacKay demonstrated a device in which the growth of silver dendrites from silver nitrate solution was controlled electrically to vary the impedances between a number of electrodes. He reported, however, that there was little change in the A.C. impedance until the dendritic connection was nearly complete (MacKay,1958).

An 'evolutionary' system utilising the electrodeposition of a metal was reported by Pask (1958). This consisted of a dish containing a concentrated, aqueous solution of ferrous sulphate. A regular array of platinum electrodes was introduced into this solution, and an external circuit was connected to these electrodes. Currents flowing

through the electrolyte caused iron fibres to grow between the electrodes. These fibres were soluble in the electrolyte and so they dissolved if insufficient current passed through them. The fibres which were used much remained whilst those which were seldom used disappeared. This, according to Pask, is analogous to the Darwinian situation in which only the 'fittest' survive.

An artificial 'synapse' which made use of the electrolytic decomposition of dilute sulphuric acid was constructed by Harmon (1958). The gases so liberated expanded a silvered rubber membrane so that it pressed against a carbon resistor and progressively shorted it.

This device had the disadvantage that the resistance could be changed in one direction only.

Widrow (1960) has produced a successful continously-variable resistor. This consists of a carbon resistor immersed in a bath containing copper sulphate solution. An additional copper electrode is introduced into the bath. By passing a direct current in the appropriate direction, copper is dissolved and deposited onto the carbon. This reduces the resistance of the carbon resistor. A current in the opposite direction will restore the resistor to its original resistance. Widrow has given the name 'memistor' (resistor with memory) to devices of this type.

Finally, Steinbuch (1961) has suggested the electro-deposition of a metal as one of the possible methods of producing the variable connections in his 'learning matrix'. The reaction suggested by him is the deposition of silver from silver bromide electrolyte.

3.(2) An electrochemical, variable-resistance device.

As mentioned in the previous chapter, the author has recently developed yet another electrochemically variable resistor which overcomes the limitations of MacKay's earlier device by growing silver dendrites in narrow channels (MacKay and Ainsworth, 1961). A number of these have now been constructed and their properties investigated. These devices appear to offer several advantages, especially from the standpoint of mass-production.

The basic design is illustrated in Figure (2.1) (Chapter II). The operation is as follows. A voltage is applied which makes electrode A positive with respect to electrode B. A dendrite forms on B and grows along the tube towards A. This reduces the resistance of the tube. The resistance at a time t, R(t) may be derived from equation (2.7),

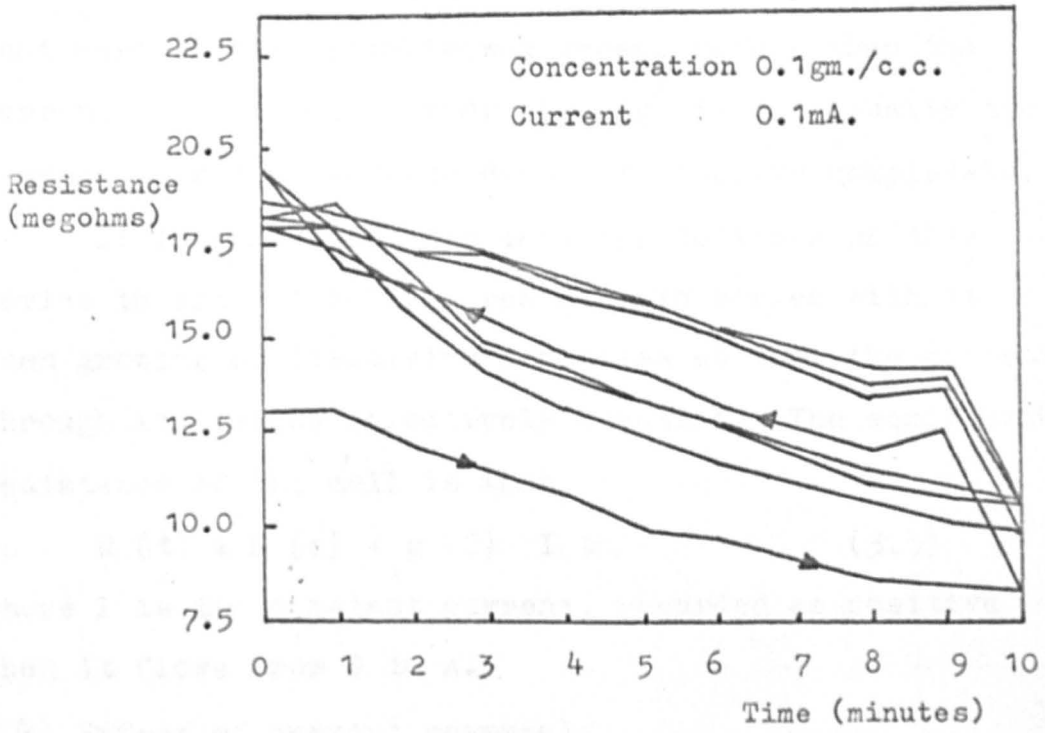
$$R(t) = R(o) - \int_o^t g(I)\ I\ dt. \quad (3.1)$$

The resistance is increased by applying a voltage in the opposite direction making A negative with respect to B. This causes the dendrite in the tube to dissolve, and silver ions to be deposited on A. As A is surrounded by a large bulk of electrolyte the current density at any point on its surface will be less than at a corresponding point on B, so there will be less tendency for a dendrite to be formed on A. If, however, one does grow, it will not much affect the impedance of the cell, as it will short circuit only the column of liquid in the wide tube, which has negligible impedance compared with that in the narrow tube.

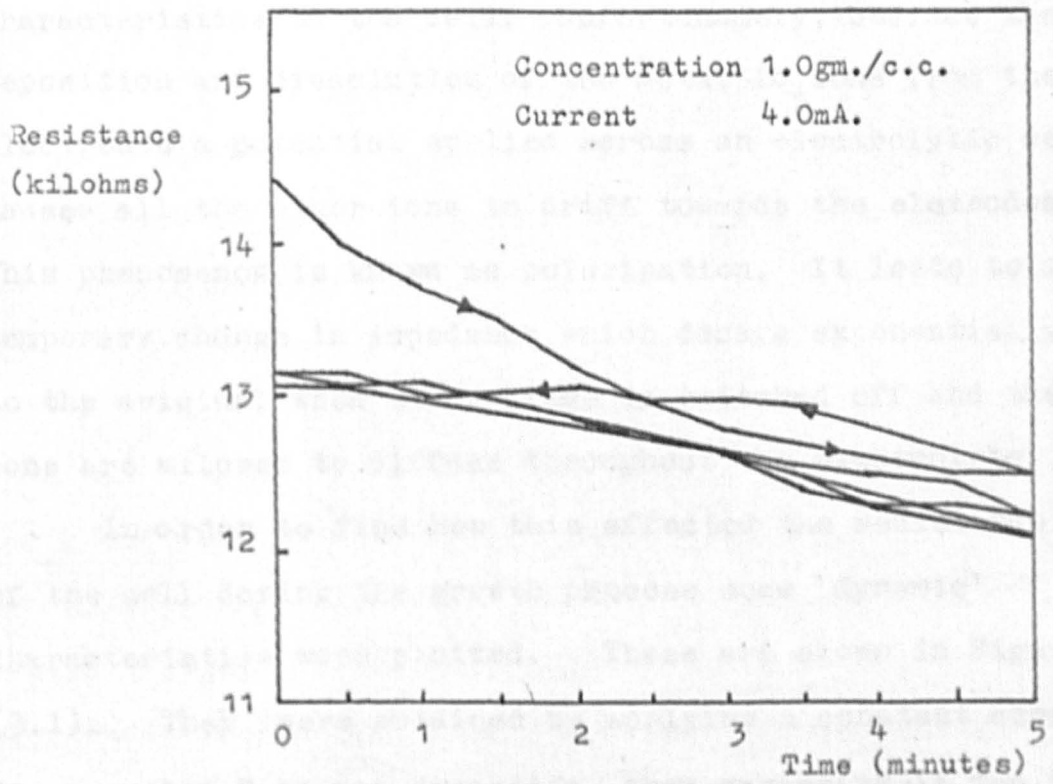
The increase in resistance of the cell is given by

$$R(t) = R(o) + \int_o^t g'(I')\ I\ dt \quad (3.2)$$

The current on which the factor g' depends is, in general, not the same as the current under the integral. Dendrites grown at different currents have different cross-sectional areas, so I' must be the current at which


Concentration 0.1gm./c.c.
Current 0.1mA.
Resistance (megohms)
22.5
20.5
17.5
15.0
12.5
10.0
7.5
0 1 2 3 4 5 6 7 8 9 10
Time (minutes)

(a)


Concentration 1.0gm./c.c.
Current 4.0mA.
Resistance (kilohms)
15
14
13
12
11
0 1 2 3 4 5
Time (minutes)

(b)

Figure 3.1.

that part of the dendrite was grown, rather than the current which is dissolving it. g' is not usually the same as g as the dendrite does not dissolve completely.

It is convenient for many applications of this device to connect a large resistor in series with it when growing or dissolving dendrites so that the current through it remains effectively constant. The equilibrium resistance of the cell is then,

$$R(t) = R(o) + g(I)\ I\ t \qquad (3.3)$$

where I is the constant current, regarded as positive when it flows from B to A.

3.(3) Effect of current reversal.

Equations (3.1) and (3.2) describe the 'static' characteristics of the cell. Unfortunately, besides the deposition and dissolution of the metallic ions from the electrodes a potential applied across an electrolytic cell causes all the other ions to drift towards the electrodes. This phenomenon is known as polarisation. It leads to a temporary change in impedance which decays exponentially to the original when the voltage is switched off and the ions are allowed to diffuse throughout the electrolyte.

In order to find how this affected the resistance of the cell during the growth process some 'dynamic' characteristics were plotted. These are shown in Figure (3.1). They were obtained by applying a constant current for a period T in one direction, then reversing it for the

same period T. The abscissae of the graphs are the instantaneous D.C. resistances of the cells. The arrows on the curves indicate the direction of the current during the first cycle.

The behaviour of the cells was similar in each case, a type of hysteresis loop being obtained. Figure (3.1.a) shows the resistance change for a cell containing an electrolyte of 0.1 gm. per c.c. silver nitrate solution when a current of 0.1 milliamps was passed in each direction for a 10 minute period. The loop is most pronounced in this case. The sudden change in resistance due to polarisation on the first current reversal is so great as almost to restore the resistance of the tube to the value it had before the dendrite was grown.

In Figure (3.1.b) another feature shows itself. The cell in this case contained a 1.0 gm. per c.c. silver nitrate electrolyte, and was driven by a 4 milliamp current for a 5 minute period. The change in resistance during the first cycle was far greater than in subsequent ones. This was due to the dendrite's not dissolving completely when the current was reversed. The small pieces which broke off during the dissolution remained in the tube and permanently lowered its resistance. The resistance per unit length of the tube was not continuously lowered, however, but remained constant after the first few cycles. After the first five cycles or so the resistance changes became

fairly constant.

From the curves of Figure (3.1) it is seen that the closest behaviour to that described by equations (3.1) and (3.2) is shown by cells containing a high concentration electrolyte, in which dendrites have been grown and dissolved by relatively large currents.

3.(4) Heating effects in tubes.

The rate at which the resistance of a cell can be changed is approximately proportional to the current through it (Figure 2.4). In order to determine the maximum rate at which resistances may be changed it is therefore necessary to determine the maximum current which may safely be passed. Preliminary experiments suggested that this was limited by the electrolyte's boiling when a certain current had been exceeded, so more detailed measurements were taken to check this.

As the internal area of the tube was so much greater than the area of its ends it was assumed that all the heat produced was conducted away through the glass wall. The rate of flow of heat from the axis of a cylindrical tube of internal radius r_1, external radius r_2, and length l is,

$$\frac{dQ}{dt} = \frac{2\pi l K (\theta_i - \theta_o)}{\log_e (r_2/r_1)} \qquad (3.4)$$

where K is the thermal conductivity of the wall,

θ_i is the temperature inside,

and θ_o is the temperature outside.

The rate of production of heat by a current I flowing through a tube of resistance R is,

$$\frac{dQ}{dt} = \frac{I^2 R}{J}$$ (J is the electrical equivalent of heat.)

It was assumed that heat was generated and dissipated only by that part of the tube which did not contain a dendrite. If this has a lengh l, then,

$$\frac{dQ}{dt} = \frac{l\, I^2}{J \pi r_1^2 \sigma} \qquad (3.5)$$

where σ is the conductivity of the electrolyte.

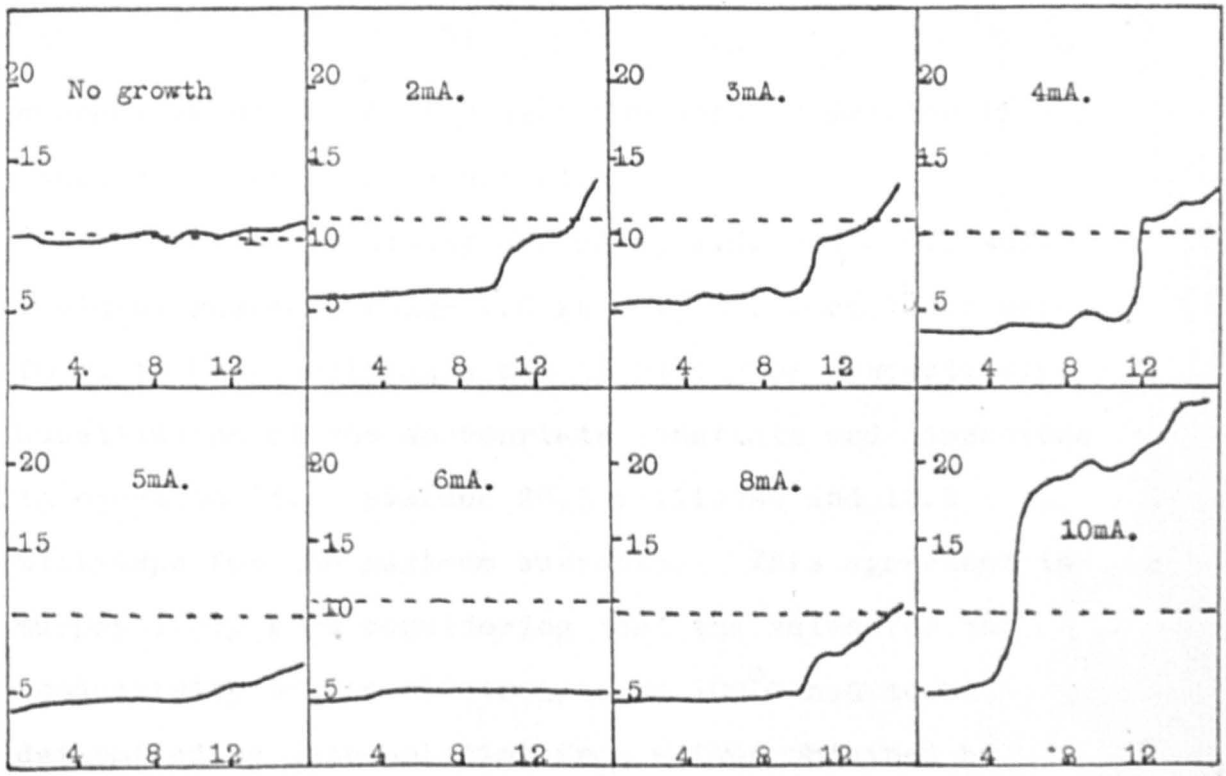
The maximum rate of loss of heat is given by equation (3.4) when θ_i is equal to the boiling point of the electrolyte. When the heat produced is equal to this the liquid will boil and the tube will fill with bubbles of vapour. The resistance of the cell will then suddenly become very large.

The maximum current which may be passed is thus,

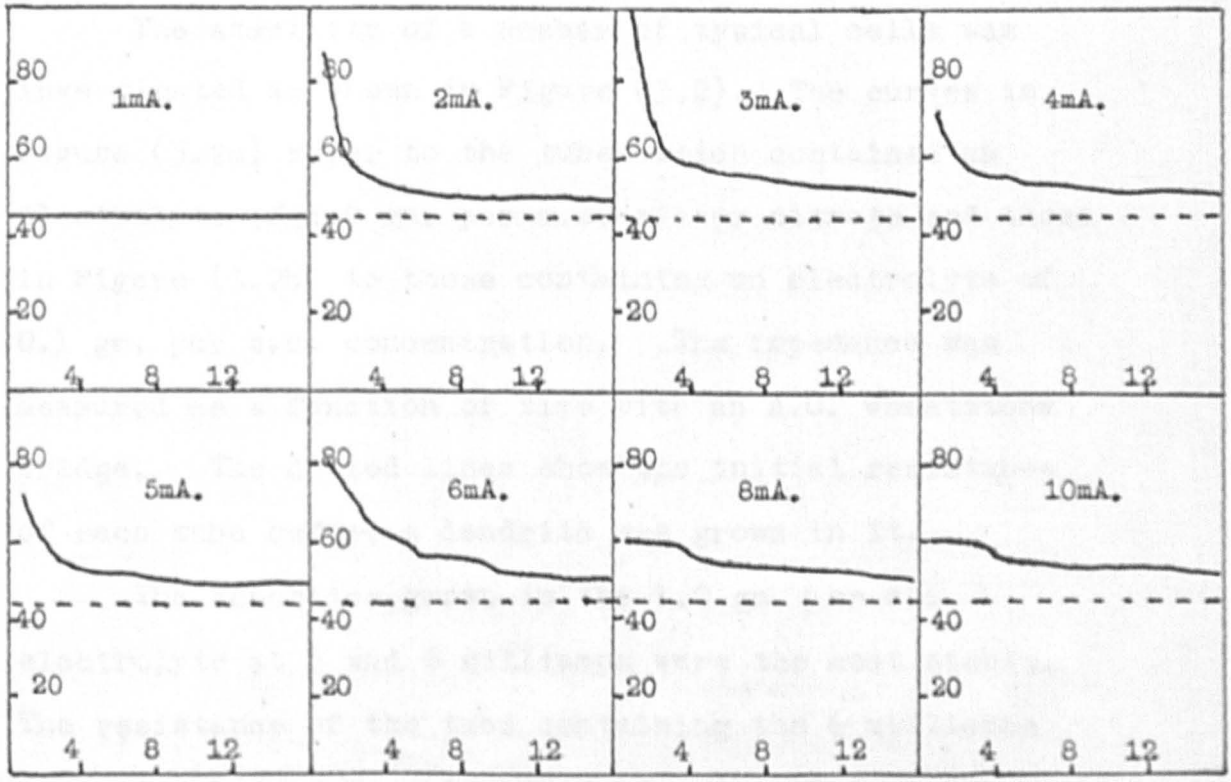
$$I_m = \left(\frac{2\pi^2 J K (\theta_m - \theta_0) r_1^2 \sigma_m}{\log_e (r_2/r_1)}\right)^{1/2} \qquad (3.6)$$

where θ_m is the boiling point of the electrolyte, and σ_m is the conductivity of the electrolyte at this temperature.

The maximum current which could be passed continuously through a tube was determined experimentally by applying an alternating voltage across a tube in which a dendrite had been grown. It was found that this current did not depend upon the length of the dendrite. Larger currents could be passed for short periods, but the maximum



(a) RESISTANCE (kilohms) versus TIME (days): CONCENTRATION 1.0gm./c.c.



(b) RESISTANCE (kilohms) versus TIME (days): CONCENTRATION 0.1gm./c.c.

Figure 3.2.

current which could be passed for periods greater than about 30 seconds was constant.

The maximum steady currents, A.C. and D.C., which could be passed through 1.0 and 0.5 mm. bore tubes were found to be 20 milliamps and 12 milliamps respectively. Substitution of the appropriate constants and dimensions in equation (3.6) yielded 20.5 milliamps and 12.6 milliamps for the maximum currents. This agreement is surprisingly good considering that the value for the conductivity of the electrolyte at 100°C had to be determined by extrapolation from values obtained by measurement at lower temperatures.

3.(5) Stability of resistance.

The stability of a number of typical cells was investigated as shown in Figure (3.2). The curves in Figure (3.2a) refer to the tubes which contained an electrolyte of 1.0 gm. per c.c. silver nitrate and those in Figure (3.2b) to those containing an electrolyte of 0.1 gm. per c.c. concentration. The impedance was measured as a function of time with an A.C. Wheatstone bridge. The dotted lines show the initial resistance of each tube before a dendrite was grown in it.

The dendrites grown in the 1.0 gm. per c.c. electrolyte at 5 and 6 milliamps were the most stable. The resistance of the tube containing the 6 milliamps dendrite changed by only 10% during the 17 days for which

the measurements were taken. After this time the cell became open-circuited. This happened because the cell was imperfectly sealed and the liquid evaporated leaving the electrode in the wide part of the tube no longer immersed in the electrolyte.

The reason why these conditions were the most stable is probably that these currents produce large metallic crystals. At lower current strengths dendritic growth is somewhat inhibited, resulting in rather erratic growth; whereas at higher current strengths the dendrites are thin and mechanically weak.

The curves obtained with the 0.1 gm. per c.c. electrolyte are peculiar in that growth caused the A.C. resistance of the tubes to increase. With this low concentration the growth must have so depleted the number of ions in the rest of the tube that there was a net increase in resistance. The resistance gradually decreased as the ions diffused along the tube but by this time the dendrites had broken up so that the resistances never fell beneath their initial values.

3.(6) Impedance of electrolyte-filled tubes.

One of the attractive properties of these devices is that their resistances can be 'read', with alternating currents without appreciably changing these resistances. If use is to be made of this property it is necessary to discover whether the impedance depends on the frequency

of the A.C. signals, and also how the A.C. impedance is related to the D.C. resistance.

Using a Wien parallel resistance bridge (v.ChapterIV) it was found that the impedance was independent of frequency over a range of 15 c.p.s. to 50 Kc. p.s. The resistive component over this range was, as far as could be ascertained, equal to the D.C. resistance. It is difficult to measure the D.C. resistance accurately as this is continually changing owing to dendritic growth and polarisation. The capacitive component of the impedance was less than 10pF for a typical cell.

The A.C. impedance of the cell appeared to be independent of the voltage across it for the range 15-900 millivolts. This was surprising as the resistance of electrolytic cells is usually non-linear in this range. Although the resistance of the electrolyte is normally independent of voltage, that of the electrode/electrolyte interface is not. In this case either the high, linear resistance of the tube was much greater than that of the interface, or the non-linearities of the interface did not manifest themselves with an A.C. signal.

3. (7) Multi-terminal devices.

It is sometimes desirable to be able to 'read' the resistance of a cell using direct current, but without changing it. This may be accomplished if a cell of the

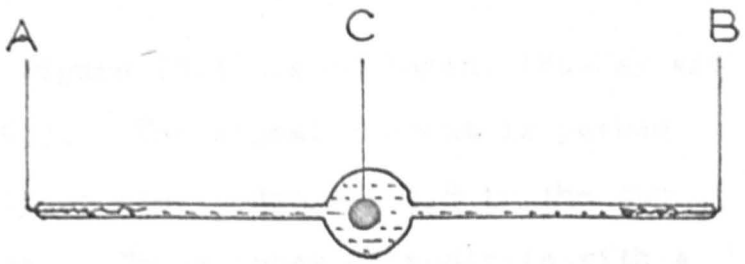


Figure 3.3.

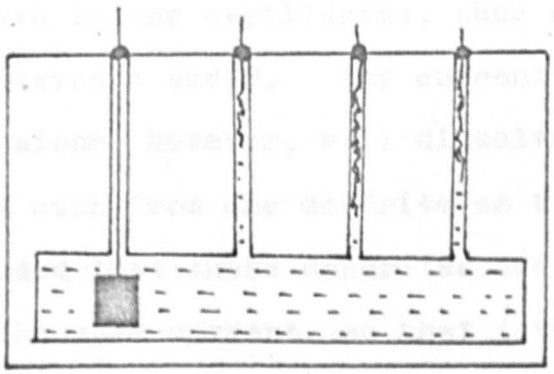


Figure 3.4.

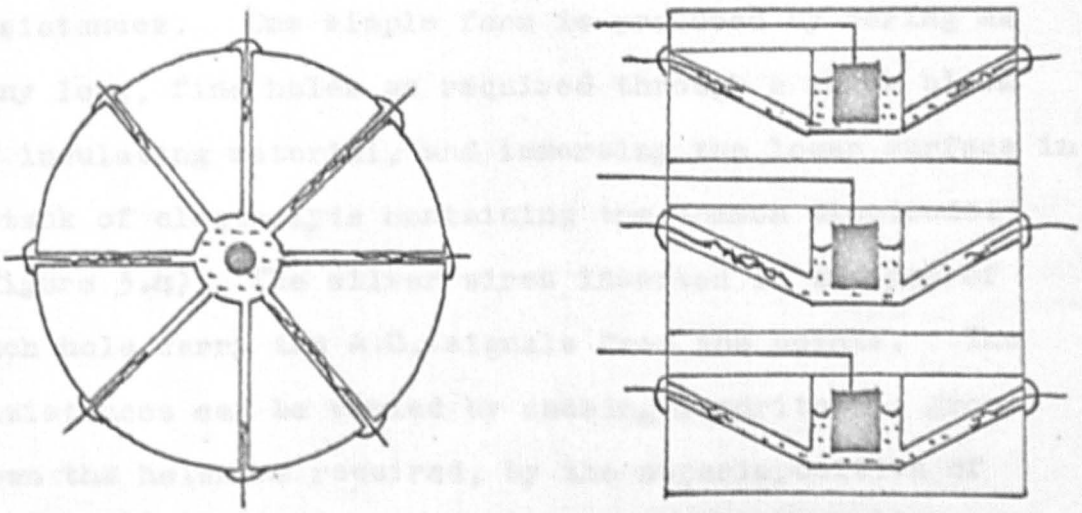


Figure 3.5.

type shown in Figure (3.3) is employed. (MacKay and Ainsworth, 1963). The signal current is passed between the silver electrodes A and B in the two capillary tubes. These tubes communicate with a common reservoir of electrolyte in which a third silver electrode C is placed. If direct current is passed between C and either or both A and B dendrites may be grown or dissolved in the capillaries, thus altering the impedance between A and B. Any current passing between A and B alone, however, will dissolve approximately as much from one dendrite as it adds to the other, provided that these dendrites are grown and dissolved with the same current, so that (over a short period) the impedance between A and B need not be affected.

This three-terminal device can be adapted and generalised to n-terminal operation, where it is desired to connect a number of points to a single terminal through variable resistances. One simple form is produced by boring as many long, fine holes as required through a thick block of insulating material, and immersing the lower surface in a tank of electrolyte containing the common electrode. (Figure 3.4) The silver wires inserted in the top of each hole carry the A.C. signals from the points. The resistances can be varied by causing dendrites to grow down the holes as required, by the superimposition of direct currents.

A radially symmetrical form which is more convenient for many purposes is illustrated in Figure (3.5). This construction makes it easy to stack large numbers of such devices vertically. (MacKay and Ainsworth, 1963).

A device of this form has been built, but as it virtually impossible to drill long, fine holes, thicker holes were drilled and glass capillary tubes were cemented into these. This gave the same internal space configuration. Dendrites were grown by passing direct currents between the peripheral and central electrodes, and it was found that the impedance of each tube could be adjusted independently.

A simpler method of manufacturing this device would be to construct a mould and fill this with molten plastic.. The holes could then be made by fine wires which could be drawn out when the plastic solidified.

3. (8) Serially-connected cells.

The maximum rate at which the resistance of a cell may be changed is limited by the current which may be safely passed, as given by equation (3.6). The advantage which is gained by employing a very narrow capillary tube (v. section 2.14) is thus reduced, as the maximum allowed current is less for narrower tubes. Also the resistance per unit length of tubes with diameters of the order of a tenth of a millimetre is rather high even when these are

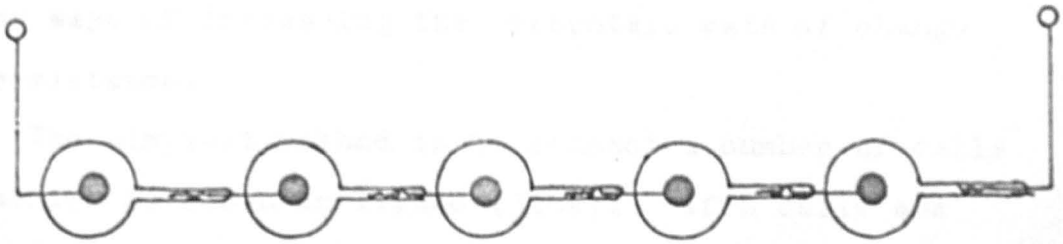


Figure 3.6(a).

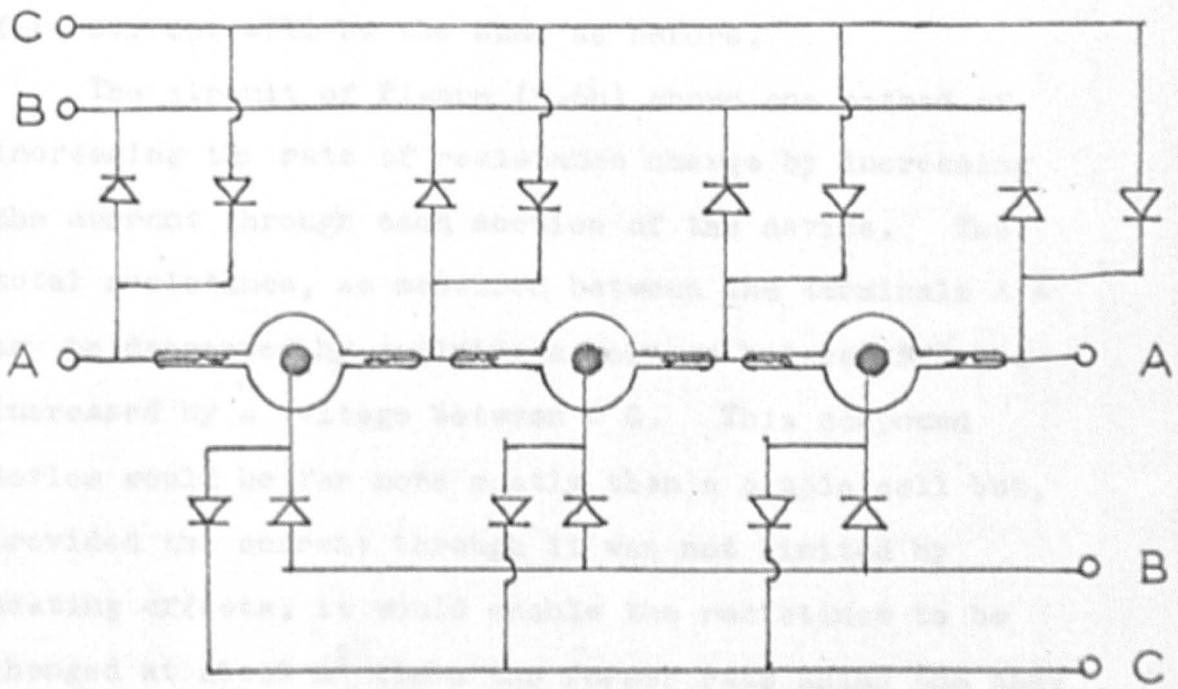


Figure 3.6(b).

filled with a saturated solution as electrolyte. So, unless high voltages are available to drive the current through these tubes, it becomes desirable to think of other ways of increasing the percentage rate of change of resistance.

The simplest method is to connect a number of cells in series as shown in Figure (3.6a). If n cells are joined in this way, the rate of resistance change will be increased n times. This arrangement, however, does not alleviate the difficulty caused by very high resistances. If each of the n cells has n^{-1} of the resistance of the cell they are replacing the total resistance, and therefore current, will be the same as before.

The circuit of Figure (3.6b) shows one method of increasing the rate of resistance change by increasing the current through each section of the device. The total resistance, as measured between the terminals A A may be decreased by applying a voltage between B B and increased by a voltage between C C. This compound device would be far more costly than a simple cell but, provided the current through it was not limited by heating effects, it would enable the resistance to be changed at about n^2 times the former rate using the same applied voltage. One of the disadvantages of this arrangement is that separate terminals are required for raising the resistance, for lowering it, and for 'sensing' it.

3. (9) Comparison of electrochemical, variable-resistance devices.

The special form of the silver-silver nitrate cell described in this chapter offers some advantages over all previous electrochemical variable-resistance devices except the 'memistor' of Widrow. Although relevant data are not available for the memistor it is probable that this would have a faster percentage rate of change of resistance than the dendrite cell. The rate of resistance change of the dendrite cell is, however, sufficiently fast for many applications, and can be increased by reducing the size of the cell,. It can also be increased many-fold by constructing the compound cells of Figure (3.6). The circuit of Figure (3.6b) can, of course, also be applied to form combinations of memistors.

For some applications the temporary resistance change caused by polarisation may make the dendrite cell unsuitable. Again the memistor should not suffer from this trouble.

It seems that if dendrites are grown at about 6 milliamps and the cells are not subjected to mechanical shocks it should be possible for the tubes to maintain their resistance values for a month or more. Probably this time would be longer if narrower capillary tubes, which would give more support to the dendrites, were employed. It is not known how this compares with the 'life-times' of other electrochemically variable resistors.

The silver nitrate cell offers a number of advantages over the memistor. The most important of these is that the range over which the resistance can be adjusted may be made as large as required simply by increasing the length of the capillary tube. Also the mean of this range can be made many orders of magnitude greater than that of the memistor. The range of the memistor is of the order of 2 to 50 ohms, which is rather low for n-terminal operation if n is greater than 10. The high resistance end of the memistor's range cannot be increased as it must be less than that of the surrounding electrolyte. On the other hand, resistances in the kilohm and megohm regions are readily obtainable with the dendrite cell.

The memistor cannot be 'read' with a direct current as this produces a potential gradient along the substrate which causes copper to dissolve from one end and deposit on the other. This causes a net change in resistance. In one modification of the dendrite cell, the three-terminal device described in section (3.7), this mode of utilisation is a possibility.

Finally, the dendrite cell possesses a feature which other variable resistors do not. It is possible to 'tailor' the kind of resistance characteristic which is required. Initially the dendrites were grown in

tubes with uniform bores because these gave a linear relationship between resistance and time at constant current. It is possible, however, to employ non-uniform bores in order to achieve other relationships, linearity between resistance and time at constant voltage for example, or a logarithmic relation between resistance and time. A whole new range of dynamic, information processing devices could be built using this principle.

REFERENCES, CHAPTER III

Drechsler, M. (1951) "Uber neue Schaltelemente und Deren Zusammenhang mit Schaltproblemen Des Zentralnervensystemes". Zeitschrift fur Naturforschung, vol. 6b, no. 7, pp.345-355.

Harmon, L.D. (1957) "An Artificial Synapse." Research Report, Bell Telephone Laboratories.

MacKay, D.M. (1958) Discussion in "Mechanisation of Thought Processes." Voll II, H.M.S.O. pp 924-5.

MacKay, D.M. and Ainsworth, W.A. (1961) "Improvements in or relating to Automatic Controlling Devices and the Like." Brit. Prov. Pat. Spec. 01288/ 61.

MacKay, D.M. and Ainsworth, W.A. (1963) "Electrolytic Growth Processes with Applications to Self-Adjusting Automata". Proceedings of D A G K Symposium, Karlsruhe, in press. (v. Appendix D of this thesis.)

Pask, G. (1958) "Physical Analogues to the Growth of a Concept" in "Mechanisation of Thought Processes". Vol. II, H.M.S.O., pp 879-928.

Steinbuch, K. (1961) "Die Lernmatrix", Kybernetik, vol. 1, no. 1, pp. 36- 45.

Widrow, B. (1960) "An Adaptive 'Adaline' Neuron using Chemical 'Memistors'." T.R. No. 1553-2, Stanford Electronics Laboratory.

CHAPTER IV

PROPERTIES OF ANODIC OXIDE FILMS

4 (1) Introduction

An alternative to the method of adjusting the impedance of an electrolytic cell by growing conducting pathways between the electrodes is by the growth of insulation on one of the electrodes (MacKay, 1961; MacKay and Ainsworth, 1961). If a film, whose resistance is much greater than that of the electrolyte, is grown on one of the electrodes, as illustrated in Figure (4.1), the impedance of the cell becomes effectively that of the film. The impedance of the cell can then be adjusted by varying the thickness of the film. This approach to the construction of a variable impedance is explored in this and the following chapter. This chapter is concerned with the methods of growth of such films and their physical properties.

4. (2) Oxide films on aluminium.

Aluminium is a metal which is, comparatively, chemically passive in air. The reason for this passivity is that the fresh metal surface rapidly develops a thin film of oxide which effectively seals it off from any further oxidation or mild chemical attack.

If, however, a piece of aluminium is made the anode of an electrolytic cell the thickness of this oxide film can be increased. There are two varieties of aluminium oxide

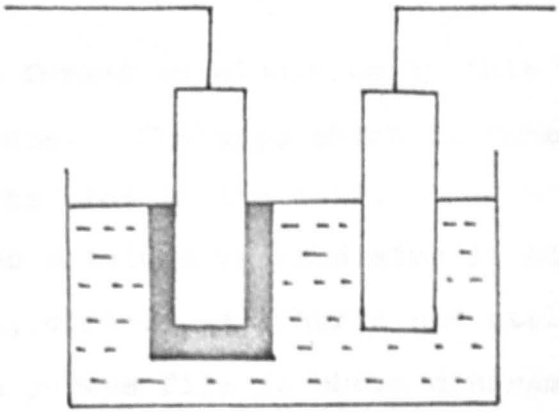


Figure 4.1.

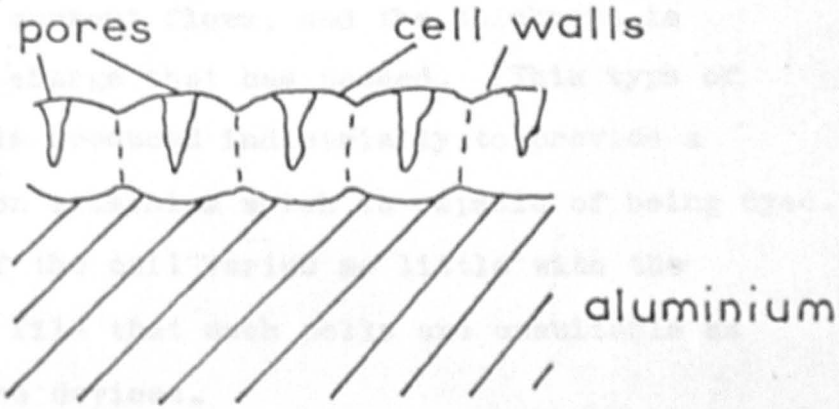


Figure 4.2.

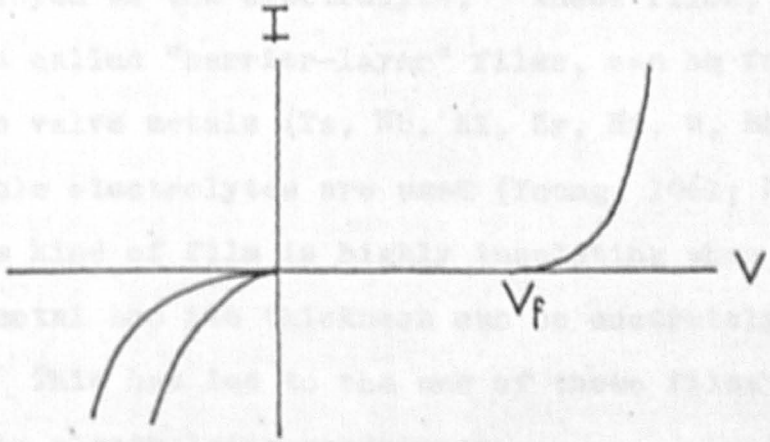


Figure 4.3.

oxide which can be formed on aluminium by this method - porous and non-porous. The type which is formed depends upon the electrolyte used in the cell.

Porous films are obtained by anodising in strong acids such as phosphoric, chromic, sulphuric and oxalic acids. The structure of a porous film is shown diagrammatically in Figure (4.2). The conductivity of the film is similar to that of the electrolyte so the film continues to grow as long as the voltage is applied to the cell. The film grows as long as current flows, and the thickness is dependent on the charge that has passed. This type of anodic oxidation is produced industrially to provide a protective film on aluminium which is capable of being dyed. The resistance of the cell varies so little with the thickness of the film that such cells are unsuitable as variable-impedance devices.

Compact, non-porous films are formed on alumimium when ammonium borate (or weak acids such as dilute sulphuric or boric) is employed as the electrolyte. These films, which are sometimes called "barrier-layer" films, can be formed on all of the valve metals (Ta, Nb, Al, Zr, Hf, W, Bi and Sn) if suitable electrolytes are used (Young, 1961; Hoar, 1959). This kind of film is highly insulating when formed on the pure metal and its thickness can be accurately controlled. This has led to the use of these films as dielectrics in electrolytic condensers.

When the first layer of film has grown the electrolyte becomes insulated from the metal. For the film to grow in these conditions the metal ions (or oxygen ions) must be transported through the oxide lattice. In general this requires a very high electric field. As the film grows under constant cell voltage the thickness of the film increases so the field strength drops. When this becomes so low that the chance of a metal ion's being able to jump the potential barrier of the lattice becomes negligible, growth effectively discontinues. Thus the film always grows to a certain thickness when a given voltage is applied for a long time. This voltage is known as the "forming voltage."

When an electrode has been formed to a voltage Vf, its current versus voltage characteristic is similar to that shown in Figure (4.3). The oxide film has a much lower resistance if the potential of the metal is made negative relative to the electrolyte. Under these conditions the voltage is sufficient to electrolyse the liquid in the cell leading to the evolution of hydrogen. As the potential of the electrode is made positive negligible current flows until the forming voltage is exceeded. The field across the oxide is then sufficient for an ionic current to flow and the behaviour becomes time dependent.

4 (3) Growth of anodic oxide films.

The first workers to make a quantitative study of the growth of non-porous anodic oxide films on aluminium were

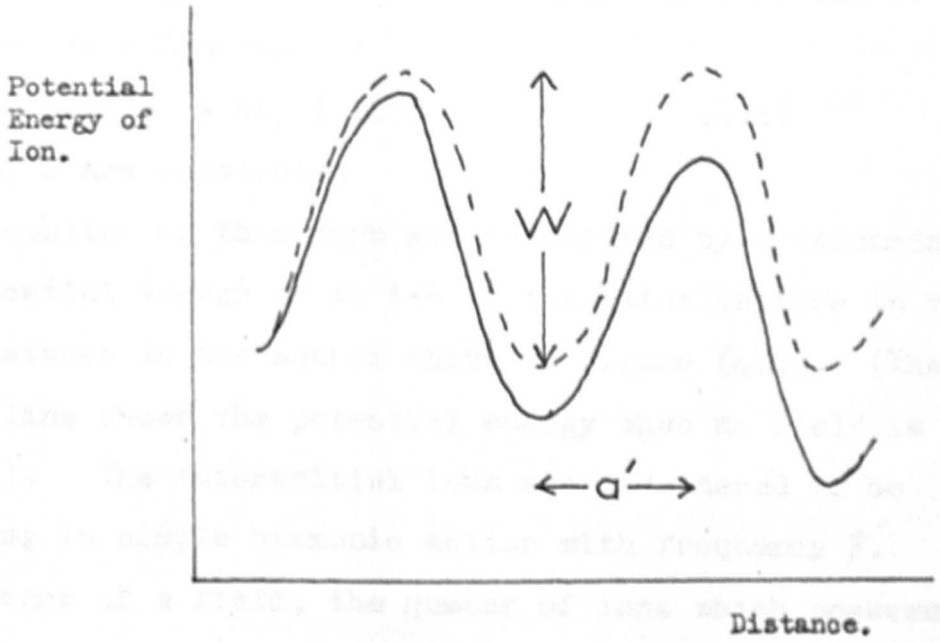


Figure 4.4.

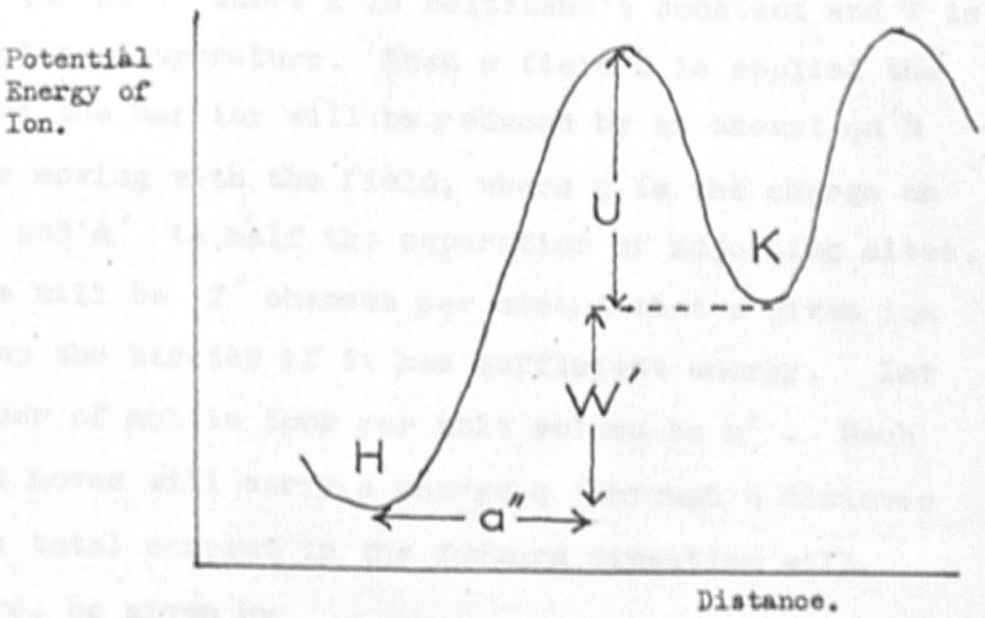


Figure 4.5.

Guntherschulze and Betz (1937). They found that the current I through a film was an exponential function of the electric field E across it

$$I = A \exp (BE) \qquad (4.1)$$

where A, B are constants.

An equation of this form may be derived by considering the potential energy of an ion in the oxide lattice to vary with distance in the manner shown in Figure (4.4). (The dotted line shows the potential energy when no field is applied). The interstitial ions are considered to be vibrating in simple harmonic motion with frequency f'. In the absence of a field, the number of ions which possess, at any given moment, sufficient energy W to jump the barrier is given by kinetic theory as proportional to $\exp - (W/ kT)$ where k is Boltzmann's constant and T is the absolute temperature. When a field E is applied the height of the barrier will be reduced by an amount $qa'E$ for ions moving with the field, where q is the charge on the ion and a' is half the separation of adjoining sites.

There will be f' chances per second that a given ion will jump the barrier if it has sufficient energy. Let the number of mobile ions per unit volume be n'. Each ion that moves will carry a charge q through a distance 2a. The total current in the forward direction will, therefore, be given by

$$I = 2a' n' q f' \exp - \left\{(W - qa' E) / kT\right\}$$

In the case of a high resistance film in order to obtain a measurable current it is necessary to apply so large a field that the current in the backward direction is negligible. The current through the film is then,

$$I = 2a'n'qf' \exp - (W/kT) \exp (qa' E/kT) \qquad (4.2)$$

which is of the same form as equation (4.1).

This theory was first applied to the growth of oxide films on aluminium by Verwey (1935).

It is not certain that the flow of ions is controlled by the barrier height in oxide films. It may be that the height of the barrier at the metal / oxide interface is the controlling factor. This case has been treated by Mott (1947) and by Cabrera and Mott (1948). The potential versus distance plot for this case is similar to that shown in Figure (4.5).

To enter the oxide an ion must jump from H to K. The probability per second that an ion will accomplish this is,

$$f'' \exp - \left\{ (W' + U - qa'' E) \ / \ kT \right\}$$

If the number of ions per unit area of electrode able to undergo transfer into the film is n", the current will be,

$$I = n''q f'' \exp - \left\{ (W' + U) \ / kT \right\} \exp (qa''E/kT) \qquad (4.3)$$

Again it is assumed that the probability of an ion jumping back is negligible. This equation also has the same form as equation (4.1).

Which of these equations applies in any given instance can only be determined by measurement and calculation of the constants A and B. This involves other difficulties. It is not known, for example, if the charge on an ion moving through a lattice, q, should be put equal to the valency of the ion. In general it seems that the Mott theory should apply to thin films and the Verwey theory to thicker ones (Young, 1961).

4 (4) Experimental work on the growth of oxide films.

Charlesby has made accurate measurements of the growth of oxide layers on aluminium. He distinguished between the current carried by the ions passing through the film (Charlesby, 1953 a) and that carried by the electrons (Charlesby, 1953 b). Film growth is only caused by the ionic current. He set out to find whether the growth of films is best described by equation (4.1) or by,

$$I_+ = A_+ E^2 \exp\left\{- B_+ / E\right\} \qquad (4.4)$$

which is what would be expected if the mechanism were similar to that of cold field emission. He found that (4.1) was the more suitable and, as he was dealing with thin films, he fitted his parameters to the theory of Mott.

The capacitance of an oxide film may be used as a measure of its thickness. It is possible to use this fact to discover when a film ceases to grow. Charlesby (1953b) found that a current still passed when the thickness of a film remained constant. He assumed that this current was

totally electronic and found that it was described by the expression,

$$I_{-} = A_{-} \sinh (B_{-} E) \qquad (4.5)$$

Charlesby derived a relation for the electron current in a similar way to that used by Mott for the ion current. The main difference in this case is that the current in the direction opposite to the field is not negligible except for extremely high fields. This led to,

$$I_{-} = 2n'' q f'' \exp - (U/kT) \sinh (a''q E/kT) \qquad (4.7)$$

Young has made a study of the growth of oxide films on tantalum and niobium. These are in many ways similar to the aluminium oxide films. Young (1954) discussed the thickness and current efficiency of the formation of oxide films on tantalum electrodes. The thickness of the film may be determined by (i) capacity measurements (ii) interference colours (films on tantalum are transparent) and (iii) determination of the quantity of charge required to form a given film, from which the mass of the oxide film may be calculated using Faraday's laws. (i) was the most convenient method of measuring thickness but it required a knowledge of the ratio of true to apparent area, and the dielectric constant. These were found using (ii) and (iii). The current efficiency of the cell was determined by measuring the amount of oxygen liberated at the electrodes. This gave the charge not used in forming the oxide.

It was found that the current efficiency approached 100% as the current density was increased.

The expression dE / d log I is known as the Tafel slope. It is predicted by both the Mott and Verwey theories to be proportional to absolute temperature. The Tafel slope was found to be substantially independent of temperature for both tantalum (Young, 1954) and niobium (Young, 1956). Dewald (1954, 1955) has shown that this result could come about if the flow of ions were controlled partially by the potential barrier at the metal / oxide interface and partially by the space charge within the bulk oxide.

4 (5) Other properties of anodic oxide films.

The thickness of oxide films on aluminium cannot be increased indefinitely by increasing the forming voltage. Forming voltages of 200 to 500 volts (depending on the purity of the aluminium) produce electroluminescence. Smith (1959) has investigated this phenomenon with a photomultiplier. He found that this effect was greater with electrodes which had only been washed than with those that had been electropolished. This suggested that defects in the lattice contributed towards electroluminescence. The high fields could more easily break down films containing many defects, and the electrons flowing through the film would collide with impurity atoms, which acted as luminescent activation, causing the emission of visible radiation.

Young (1954) has discussed the photo-effects in tantalum oxide films in ultra-violet light. The photocurrent was measured as a function of field and film thickness. At a given field he found that the photocurrent increased rapidly with thickness up to a few hundred Ångstroms, and then remained constant. This suggested strong absorption in the oxide.

An effect which may be due to space charge has been reported by van Geel and Pistorius (1956). They charged an electrolytic condenser and then discharged it. Some time later they found that the condenser once more had a voltage across its plates. This voltage, which they refer to as the "residual voltage", was found to reach a maximum after about 5 minutes, and then slowly decrease. They explained this phenomenon by assuming that the oxide lattice contains closely packed oxygen ions. A unit cell consists of 32 oxygen ions and sites for 96 aluminium ions. On average there are $21\frac{1}{3}$ aluminium ions statistically distributed amongst these sites. When the charging voltage is applied the aluminium ions are displaced to adjacent sites. After this external voltage has been removed the ions return to their original distribution by diffusion, and in so doing recharge the condenser.

4 (6) The breakdown of aluminium oxide films.

The forming of an oxide film on an aluminium anode causes the resistance of the cell to be increased. It is

also possible to decrease this resistance by reversing the polarity of the cell so that the current flows through the film in the low resistance direction.

A cell consisting of an aluminium electrode, a platinum electrode and an electrolyte of aqueous ammonium borate was built. A film was formed on the aluminium by making this the anode of the cell, The resistance of the cell was noted. The current was then reversed for a short time, and then switched back to its original direction. A large current flowed which gradually reduced until the cell regained its original resistance. The time constant of this process was very much greater than would have been expected if the cell were merely acting as a condenser and being recharged. It was assumed that the reverse current had in some way broken the film and reduced its resistance. This phenomenon will be referred to as "erosion" of the film.

In order to study the erosion of anodic oxide films two methods of measuring the resistance of the film immediately after breakdown were devised. It is not possible to measure this directly as applying a forward voltage causes the film to grow, changing the resistance of the cell. (The resistance changes most rapidly immediately after the voltage is applied). Nor can the resistance be measured simply by applying an alternating current as this too would produce growth and breakdown of the film and it is not likely that these two processes would cancel each other

because of the rectifying properties of the film (Figure 4.3). The resistance of the film can be measured, however, by direct current using an extrapolation method, and by alternating current if a cell of special design is employed.

4 (7) Extrapolation method.

The relation between current and field strength in an anodic oxide film is given by equation (4.1). If x is the thickness of a film and V is the voltage across it, equation (4.1) becomes,

$$I = A \exp BV/x \qquad (4.8)$$

Young (1954) found that in the formation of oxide films on tantalum the current efficiency was about 100%. Assuming that this is also true for aluminium, the thickness of the film will be approximately proportional to the charge which has passed.

$$x = C \int_0^t I dt \qquad (4.9)$$

where C is a constant.

Assuming that the resistance of the film is much greater than that of the external circuit so that the voltage across the cell remains constant, equation (4.8) becomes,

$$x = \frac{B\ V}{\log_e (I/A)} \qquad (4.10)$$

Combining (4.9) and(4.10)

$$\frac{BV}{C} = \log_e (I/A) \int_0^t I dt$$

Differentiating with respect to t, this becomes,

$$t = -\frac{BV}{C} \int \frac{dI}{[I \log_e (I/A)]^2} \qquad (4.11)$$

For the the range of I used in the following experiments a 500% change in I causes only about a 50% change in loge (I/A). As equation (4.11) is only to be used for extrapolation it is a reasonable approximation to regard loge (I/A) as constant. Equation (4.11) then simplifies to,

$$t = -D \int \frac{dI}{I^2} \qquad (4.12)$$

where $D = \frac{BV}{C[\log_e (I/A)]^2}$ is a constant.

Hence $$I (t + t_o) = D \qquad (4.13).$$

where t_o is the constant of integration.

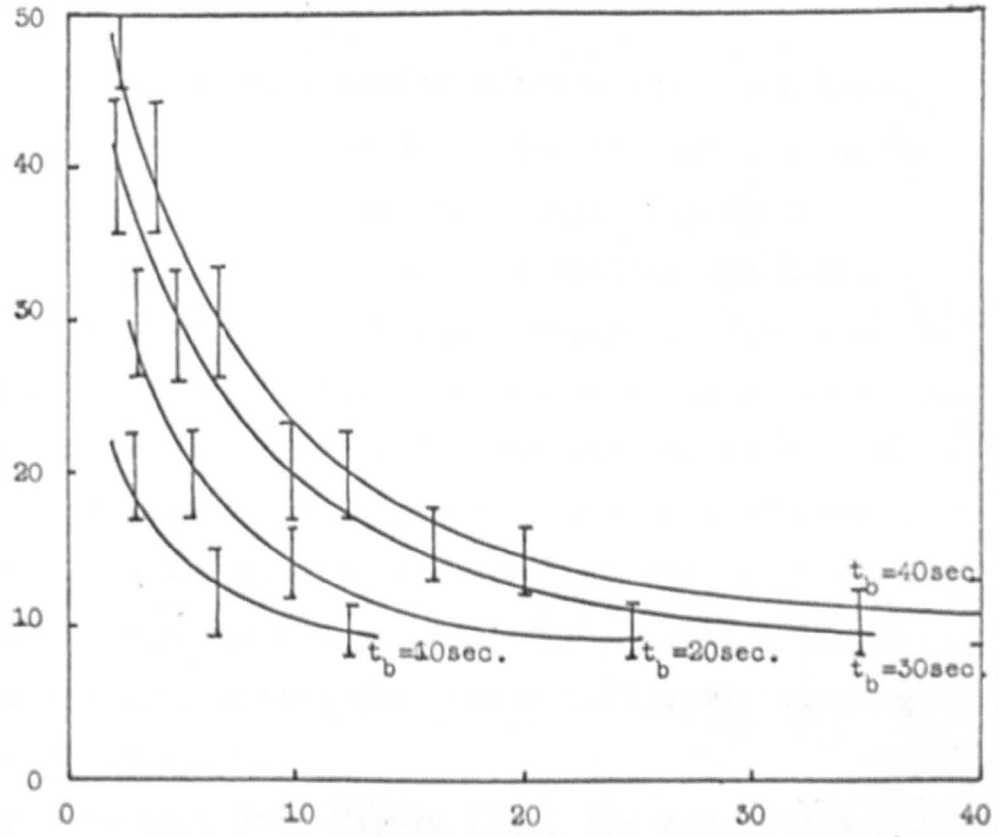
In the derivation of equation (4.13) it was assumed that all the current passing through the film caused film growth but, as has been shown by Charlesby (1953b), a residual current flows when the film has ceased growing. This residual current may be allowed for by modifying equation (4.13) to,

$$(I - I_o) (t + t_o) = D \qquad (4.14)$$

I_o may be a function of I, but as it is only significant when I approaches zero, it is effectively a constant.

4 (8) Results of the extrapolation method.

A film was formed at 24 volts on an aluminium



CURRENT (microamps) versus TIME (seconds), t_f.

Figure 4.6.

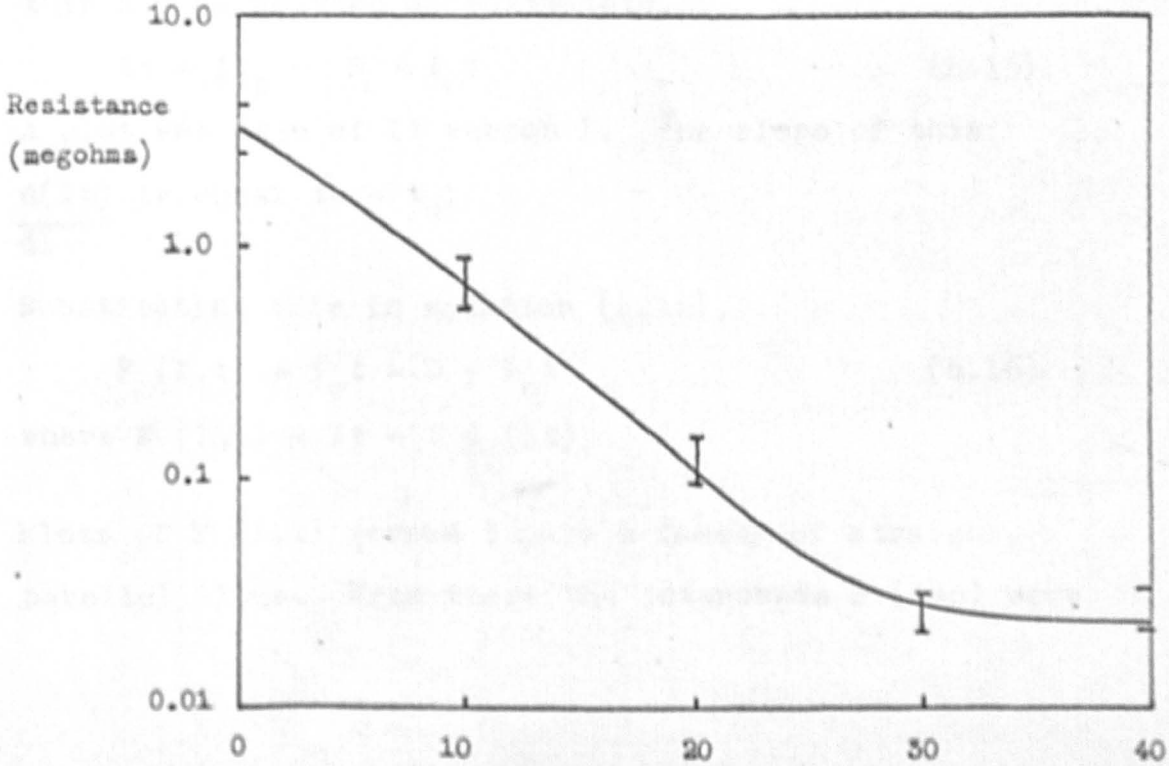


Figure 4.7.

electrode in an ammonium borate electrolyte, then broken down by reversing the polarity of the battery across the cell. This was done several times until the cycle became consistent. The voltage was applied until the resistance of the film was about 6 megohms. The polarity was reversed for a time t_b, then the original polarity was restored and the times taken for the current to drop to certain values noted. This was done for each of several values of t_f. The results are shown in Figure (4.6). The scatter of the results was caused by uncertainties in reading a moving pointer, and possibly slightly varying initial conditions.

As can be seen from Figure (4.6) the evaluation of I for t = 0 presents some difficulty. This was overcome by the use of equation (4.14). As t tends to zero this may be written approximately,

$$It = It_o = D + I_o t_o \qquad (4.15).$$

A plot was made of It versus I. The slope of this $\frac{d(It)}{dI}$ is equal to $- t_o$;

Substituting this in equation (4.14),

$$F(I,t) + I_o t = D - I_o t \qquad (4.16)$$

where $F(I,t) = It - I \frac{d}{dt}(It)$

Plots of F (I,t) versus t gave a family of straight, parellel lines. From these the intercepts F (t=o) were

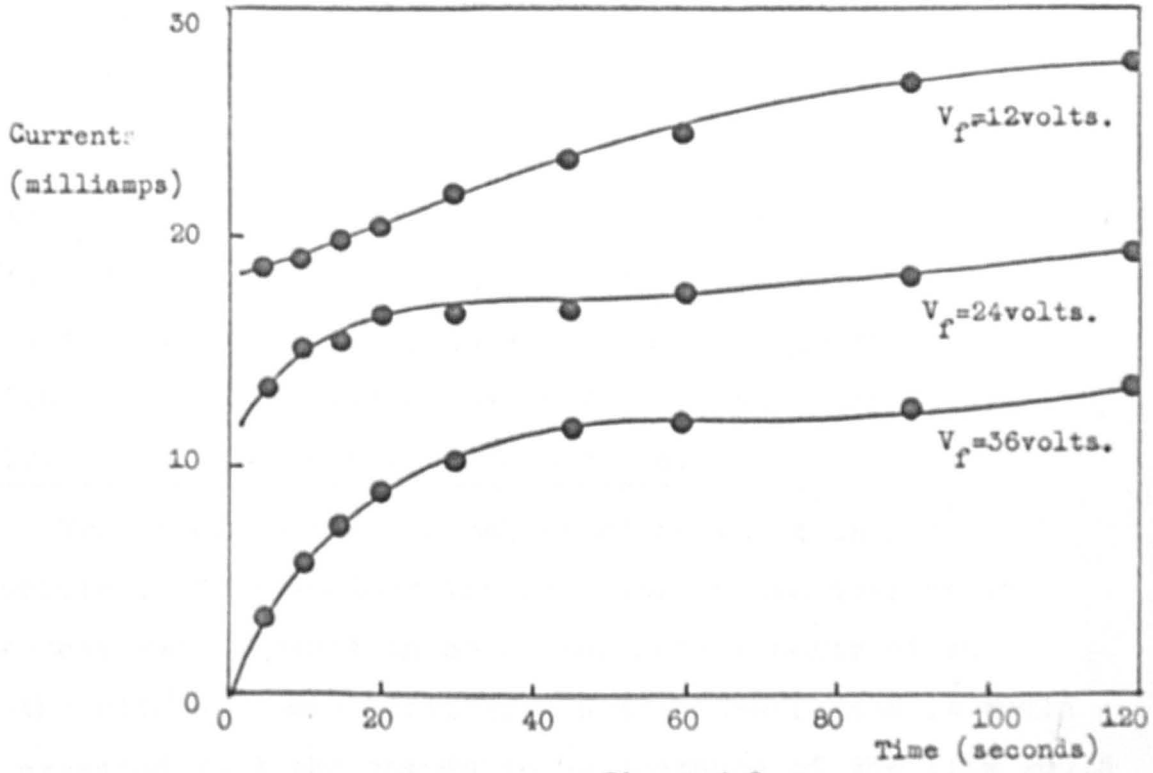


Figure 4.8.

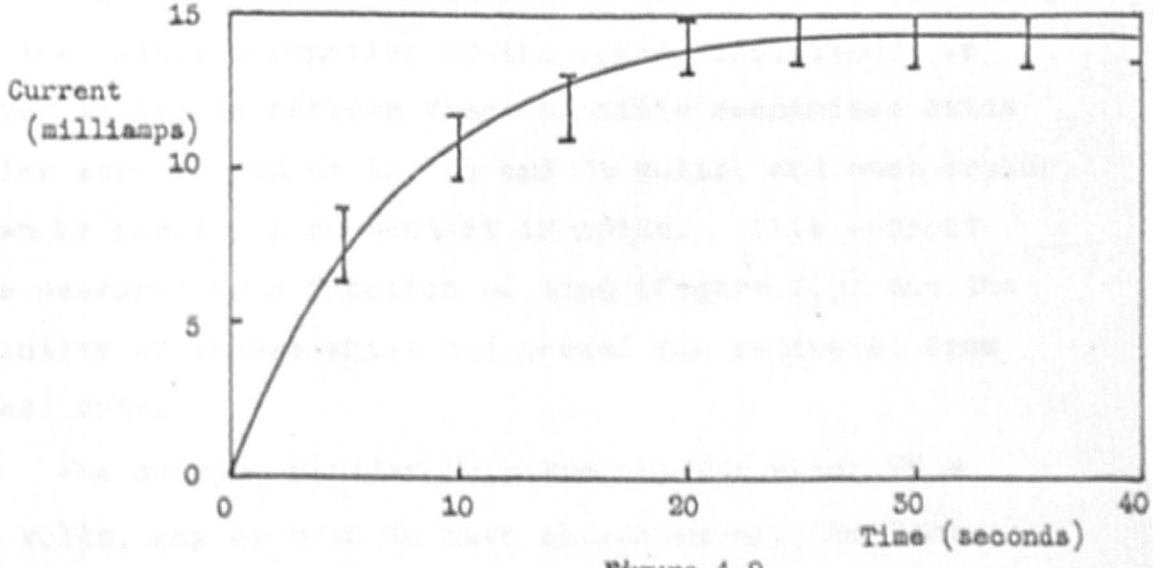


Figure 4.9.

V_f (volts)	Resistance (kilohms)	Charge (millicoulombs)	Energy (joules)
36	55	1630	9.8
24	56	1700	20.5
12	56	1710	31.0

Table 4.1.

obtained, and I(t=o) was calculated.

The value of the resistance R of the film immediately after breakdown was obtained from these data. Figure (4.7) shows that R, the forward resistance of a typical oxide film, decreased as t_b, the time for which the applied voltage was reversed, was increased.

4 (9) Nature of the breakdown process.

From the experiments described above it is not possible to tell whether the film was broken down by an electrochemical reaction or by mechanical destruction of the film by ohmic heating. In the former case it would be expected that the change in conductance of the film would be a function of the quantity of charge which passed, and in the latter a function of the power dissipated. In order to decide between these possible mechanisms oxide films were formed at 12, 24 and 36 volts, and each broken down by passing a current at 12 volts. This current was measured as a function of time (Figure 4.8) and the quantity of charge which had passed was estimated from these data.

The curves, particularly the one for which Vf = 36 volts, may be seen to have approximately the form,

$$I = X - Y \exp(-bt) \qquad (4.17)$$

at least where t is fairly small. The 36-volt data are shown on a larger scale in Figure (4.9), fitted to a curve of this shape with X = Y = 13.5, b = 0.18.

The total charge passed and the total energy dissipated in the different films were calculated, and these are shown in Table (4.1). The forward resistance of the cell immediately after the original polarity had been restored was measured by the extrapolation method described above. This resistance after breakdown is also shown in Table (4.1). The values of resistance are about the same in each case because the reverse currents were passed for different times.

It can be seen that the amount of damage caused by reversing the polarity of the cell is much more strongly related to the quantity of charge passed than the heat produced. This suggests that the erosion of the film is an electrochemical process.

4 (10) The A.C. impedance of anodic oxide films.

The impedance of an anodic oxide film may be determined using an A.C. signal provided that the mean level of the signal is biassed so that the aluminium electrode is never negative relative to the platinum one. Also it must never be more positive than the forming voltage or the film will grow and the impedance will change.

As an alternative to biassing the A.C. signal a cell of the type shown in Figure (4.10) may be employed. The D.C. voltage used to grow the film is applied between the platinum electrode A and the aluminium electrodes B and C.

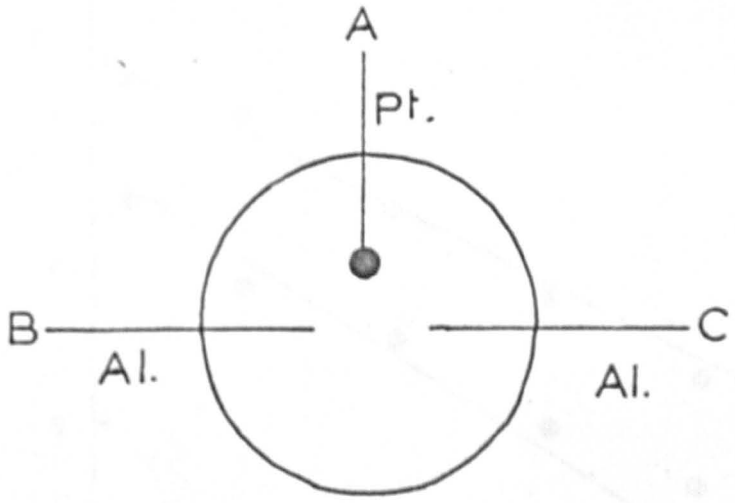
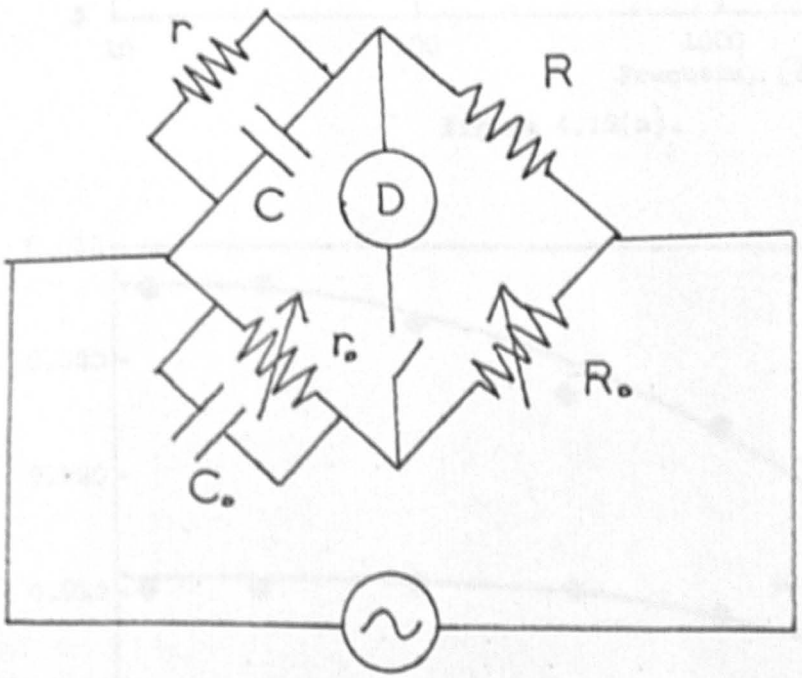


Figure 4.10.



$$r = \frac{r_o R}{R_o} \qquad C = \frac{C_o R_o}{R}$$

Figure 4.11.

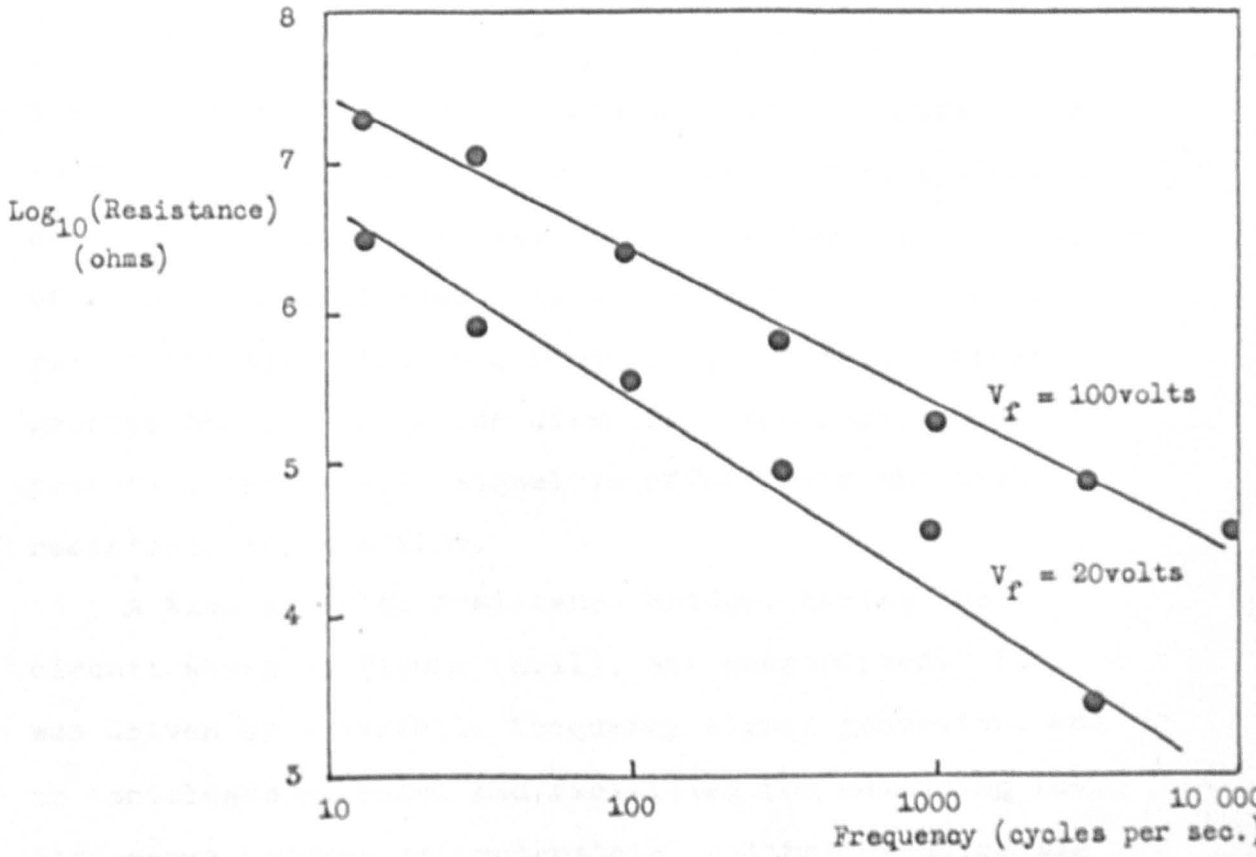


Figure 4.12(a).

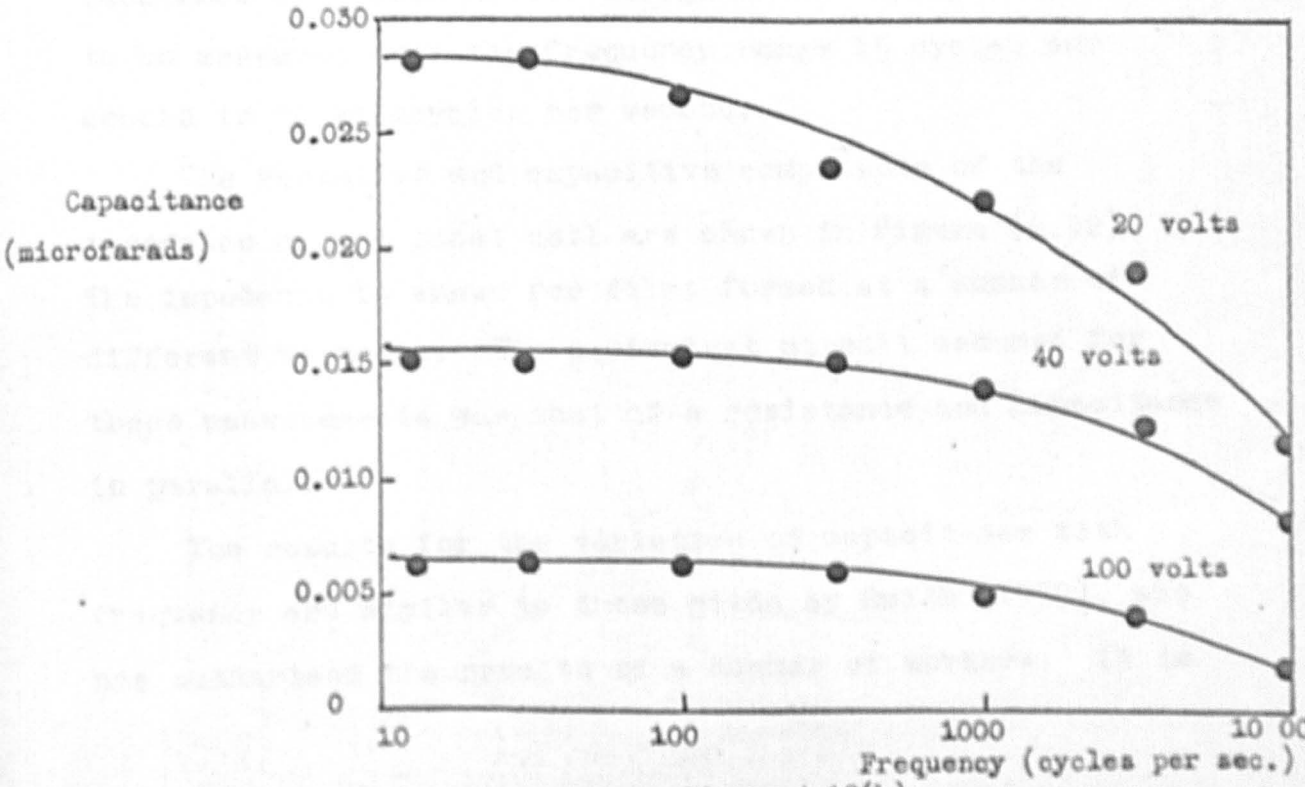


Figure 4.12(b).

The A.C. signal is applied between B and C. During one half of the cycle the high resistance of B will limit the current and during the next half cycle the high resistance of C. will predominate. As the resistance in the high resistance direction is a hundred to a thousand times greater than in the other direction, the resistance presented to the A.C. signal is effectively the high resistance all the time.

A Wien parallel resistance bridge, having the circuit shown in Figure (4.11), was constructed. It was driven by a variable frequency signal generator, and an oscilloscope, which had facilities for measuring the difference between two potentials, neither of which was earthed, was used as the detector. This enabled the impedance of a cell of the design shown in Figure (4.10) to be measured over the frequency range 15 cycles per second to 50 kilocycles per second.

The resistive and capacitive components of the impedance of a typical cell are shown in Figure (4.12). The impedance is shown for films formed at a number of different voltages. The equivalent circuit assumed for these measurements was that of a resistance and capacitance in parallel.

The results for the variation of capacitance with frequency are similar to those given by Smith (1959), who has summarised the results of a number of workers. It is

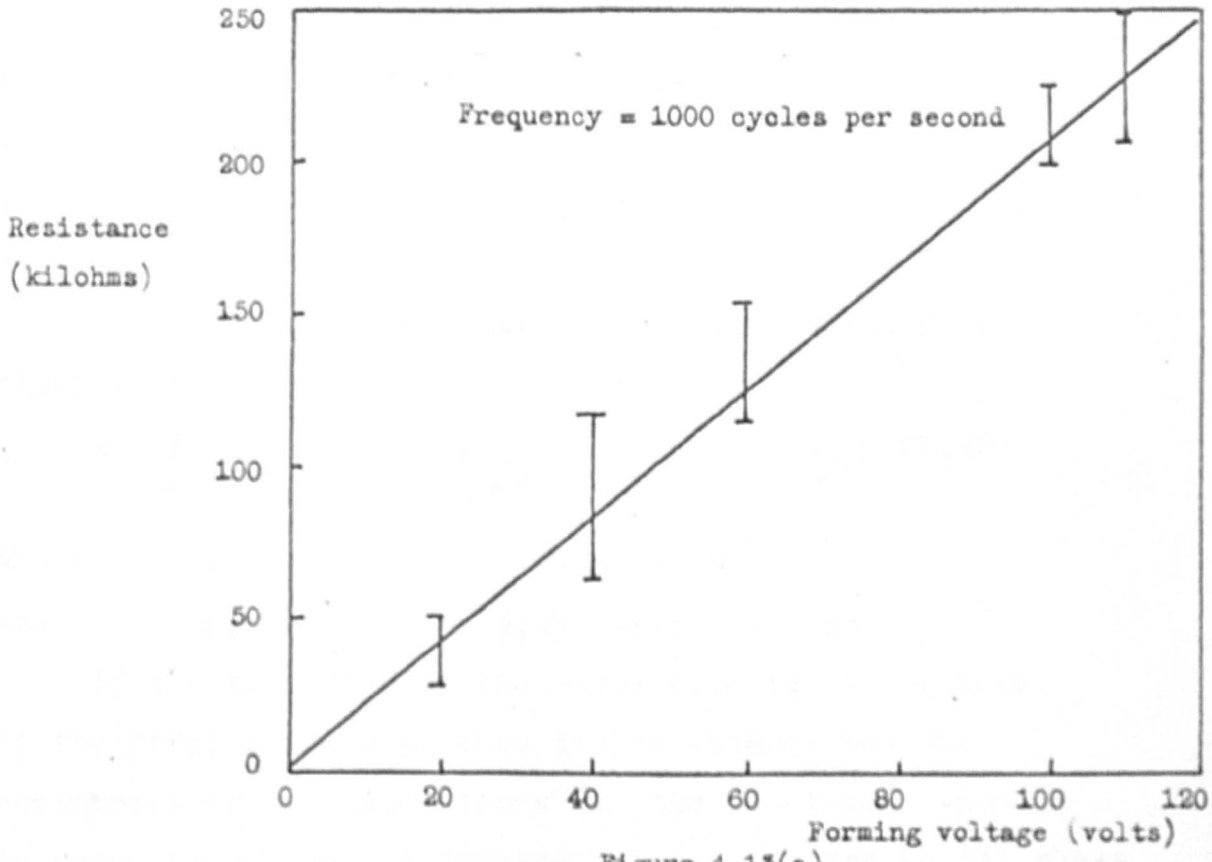


Figure 4.13(a).

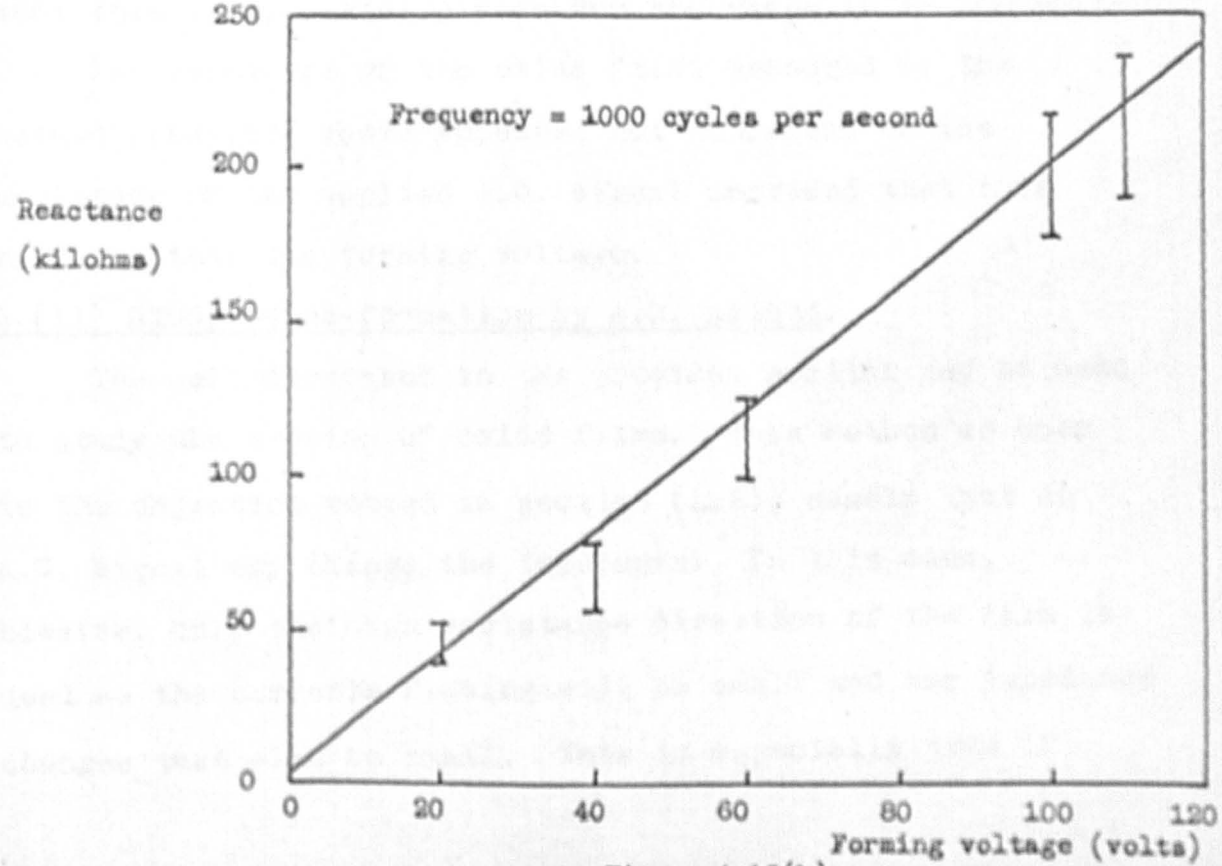


Figure 4.13(b).

interesting to note that the resistance is approximately inversely proportional to the frequency. A similar result was obtained by Young (1955) for oxide films on niobium.

The resistance and capacitance of an ideal film of thickness ℓ and area a may be written as,

$$R = \frac{\ell}{\sigma a} \quad ; \quad C = \frac{\epsilon a}{4\pi \ell} \qquad (4.18)$$

where σ is the conductivity,

and ϵ is the dielectric constant.

If the thickness of the oxide film is proportional to the forming voltage, then the resistance and the reciprocal of the capacitance (or the reactance) should be proportional to the forming voltage. Figure (4.13) shows that this is approximately so for the range 10 to 100 volts.

The impedance of the oxide films measured by the method described above appeared not to depend on the amplitude of the applied A.C. signal provided that this was less than the forming voltage.

4 (11) Study of de-formation by A.C. method.

The cell described in the previous section may be used to study the erosion of oxide films. This method is open to the objection voiced in section (4.6), namely that an A.C. signal may change the impedance. In this case, however, only the high resistance direction of the film is used so the currents flowing will be small and any impedance changes must also be small. This is especially true if

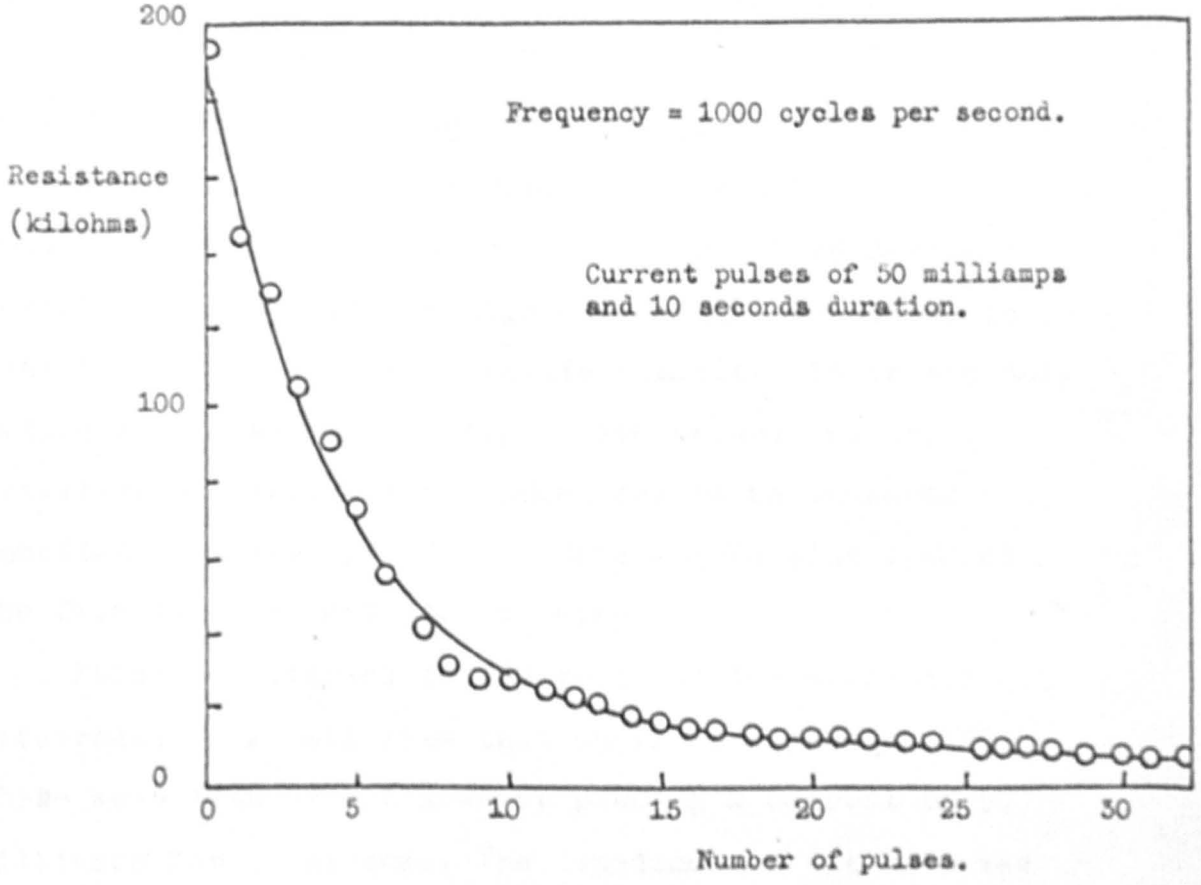


Figure 4.14(a).

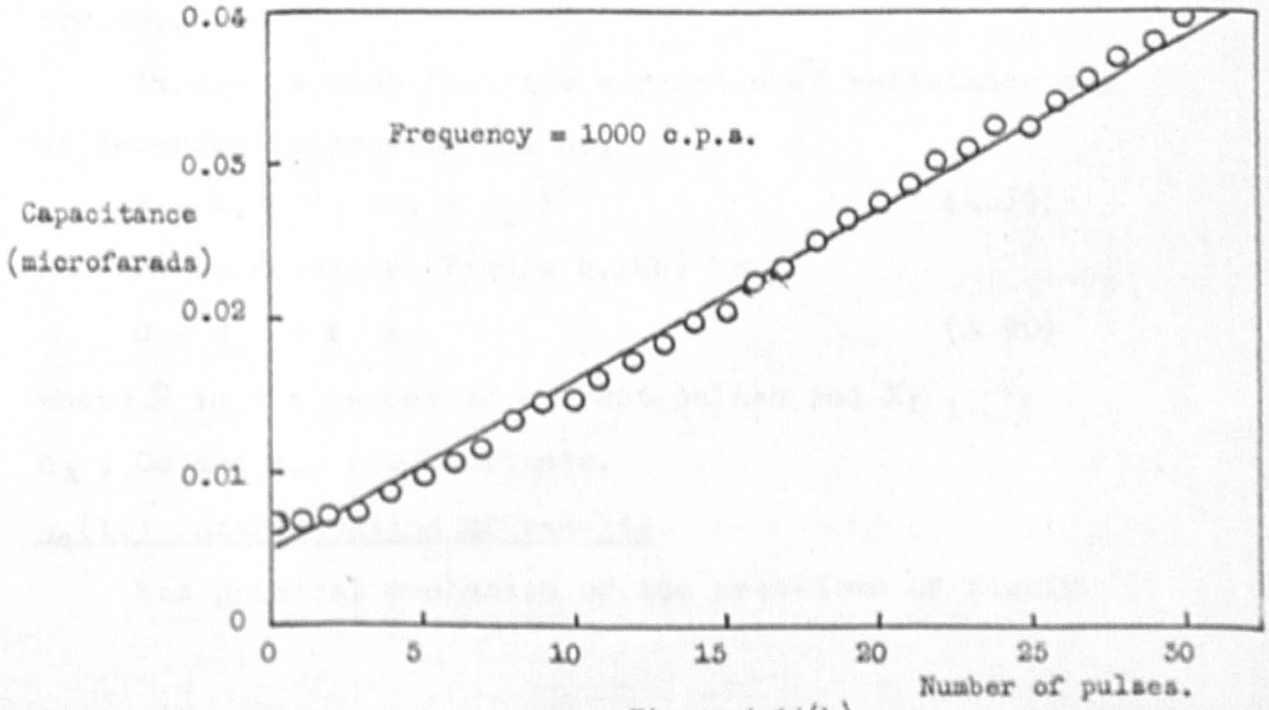


Figure 4.14(b).

only very low voltage signals are applied. Also as this cell is symmetrical it is possible that any change brought about during one half cycle will be nullified during the next. The main justification for using this method is that it gave fairly reproducible results. It is the only method which enables the capacitive as well as the resistive component of the impedance to be measured immediately after breakdown. This method also enabled the film to be eroded step by step.

Films were formed to 100 volts on the aluminium electrodes of a cell like that shown in Figure (4.10). These were then broken down by passing a current of 50 milliamps for 10 seconds. The impedance of the film was measured, then the current was passed for a further 10 seconds. Figure (4.14) shows how the resistance and capacitance of the film changed at each stage during this process.

It can be seen that the variation of resistance may be described approximately by,

$$R = X_1 + Y_1 \exp - b_2 \bar{N} \qquad (4.19)$$

and the capacitance (Figure 4.14b) by,

$$C = C_o + k_1 \bar{N} \qquad (4.20)$$

where $\bar{N}$ is the number of current pulses and X_1 , Y_1 , b_2 , Co and k_1 are constants.

4.(12) Interpretation of results.

The physical mechanism of the breakdown of highly

insulating oxide films must be essentially different from that of the formation, even though the electrochemistry may be similar. Electric charge flows most easily through those parts of the film which have least resistance. If the current is ionic it will change the impedance of those parts of the film through which it flows. During the formation of a film those areas which have less resistance will have a higher current density, and this will increase their resistance. If a film is homogeneous it should therefore grow to a uniform thickness.

During the erosion those areas of the film which are less resistive will again have a higher current density but, if dissolution of the film is taking place, this will tend to decrease the resistance of these areas still further. This will lead to a further increase in current and an even greater drop in resistance. The majority of the current will rapidly be conducted by a number of 'pits' which have been formed right through the film. This is shown schematically in Figure (4.15).

If a few simplifying assumptions are made it is possible to deduce from this model the form of the variation of impedance with time as the film is eroded.

Suppose that an insulating film, of area a and initially of uniform thickness l, is in contact with an electrolyte. A voltage V is applied in such a direction as to cause the film to dissolve. If the film were perfectly

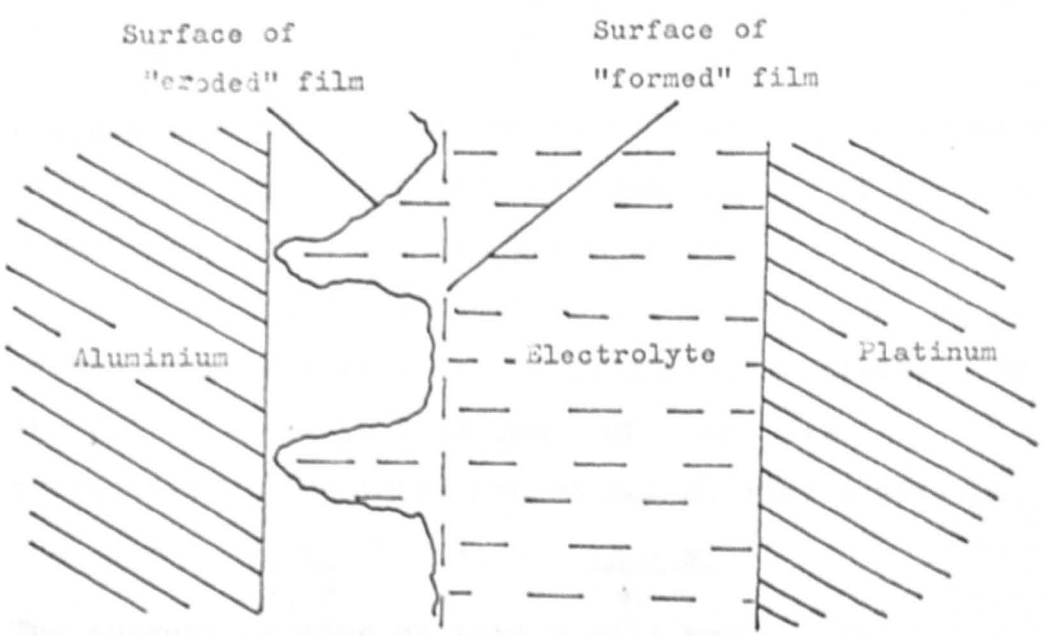


Figure 4.15.

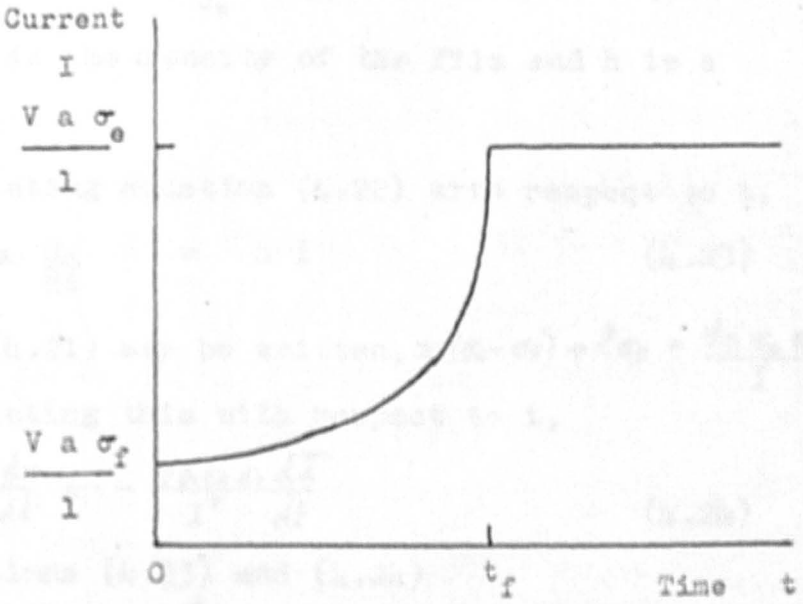


Figure 4.16.

homogeneous and uniform the current density would be the same in all parts and the thickness of the film would remain uniform during the dissolution.

Consider first this ideal case. Suppose the thickness of the film at time t is x. If the conductivities of the film and electrolyte are σ_f and σ_e respectively, the total resistance at time t will be,

$$\frac{x}{\sigma_f a} + \frac{l - x}{\sigma_e a}$$

The current passing at time t will be,

$$I(t) = \frac{V a \sigma_e \sigma_f}{x(\sigma_e - \sigma_f) + l\sigma_f} \qquad (4.21)$$

Now, assume that the mass of film dissolved is proportional to the charge passed (Faraday's law)

$$\rho a (1-x) = h \int_0^t I dt \qquad (4.22)$$

where ρ is the density of the film and h is a constant.

Differentiating equation (4.22) with respect to t,

$$-\rho a \frac{dx}{dt} = h I \qquad (4.23)$$

Equation (4.21) may be written, $x(\sigma_e - \sigma_f) + l\sigma_f = \frac{V a \sigma_e \sigma_f}{I}$.

Differentiating this with respect to t,

$$(\sigma_e - \sigma_f)\frac{dx}{dt} = -\frac{V a \sigma_e \sigma_f}{I^2}\frac{dI}{dt}. \qquad (4.24)$$

From equations (4.23) and (4.24)

$$\int \frac{dI}{I^3} = \alpha \int dt, \qquad (4.25)$$

where $$\alpha = \frac{\rho a^2 V \sigma_e \sigma_f}{(\sigma_e - \sigma_f) h}.$$

Integrating equation (4.25),

$$\beta - \frac{2}{I^2} = \alpha t \quad , \qquad (4.26)$$

where β is the constant of integration.

When t = 0, x = 1, $I = \frac{V a \sigma_f}{l}$

$$\therefore \beta = 2\left(\frac{l}{V a \sigma_f}\right)^2$$

When x = 0, $I = \frac{V a \sigma_e}{l}$

This occurs when $t = t_f = \frac{2l^2}{V^2 a^2 \alpha}\left(\frac{1}{\sigma_f^2} - \frac{1}{\sigma_e^2}\right)$ (4.27)

Between the times t = 0 and $t = t_f$ the current is given by equation (4.26). For t greater than t_f the film will have dissolved completely so the current will remain constant. This is shown in Figure (4.16).

Now consider a real film in which the thickness varies slightly from point to point. This may be considered divided into N equal elementary parts each of which is small enough to be considered of uniform thickness.

It can be seen from Figure (4.16) that most of the change in current passing through a film of uniform thickness occurs over a period of time short compared with t_f. As an approximation, therefore, it is possible to express the variation of current with time by the step function,

$$I = \frac{V a \sigma_f}{l} \quad , \quad t < t_f$$
$$I = \frac{V a \sigma_e}{l} \quad , \quad t \geq t_f \quad . \qquad (4.28).$$

This is equivalent to assuming that on any given elementary part the film remains intact until a certain charge has passed, after which it is removed suddenly and completely.

Suppose at time t, n of these N parts are covered by film, and that the film has been removed from (N - n) of these and has been replaced by electrolyte. The total current passing will be,

$$I(t) = Ie + I_f$$

$$= \frac{Va}{Nl}\left\{ \sigma_e (N-n) + \sigma_f n \right\} \qquad (4.29)$$

as each elementary part will be of area a/N.

In the time inverval t to t + dt a total charge $I_f dt$ will have passed through the film. The charge required to dissolve the film completely from one elementary part can be calculated from Figure (4.17) or approximately from equation (4.27), giving,

$$q = \frac{2 l \sigma_f N}{Va\alpha} \left(\frac{1}{\sigma_f^2} - \frac{1}{\sigma_e^2} \right)$$

The average number of elementary parts of film dissolved in a time dt will be,

$$dn = - \frac{I_f\, dt}{q} \qquad (4.30)$$

where I_f is the current through the film during this period. (The negative sign is introduced because the number of parts covered by film n is reduced).

Substituting for I_f, equation (4.30), becomes,

$$\frac{dn}{dt} = - \frac{V a \sigma_f n}{N q l} \qquad (4.31)$$

Rearranging equation (4.29),

$$n = \frac{N}{(\sigma_e - \sigma_f)} \left(\sigma_e - \frac{Il}{Va} \right) \qquad (4.32)$$

Differentiating with respect to t,

$$\frac{dn}{dt} = - \frac{Nl}{(\sigma_e - \sigma_f) Va} \cdot \frac{dI}{dt} \qquad (4.33)$$

Substituting (4.32) and (4.33) in equation (4.31),

$$\frac{dI}{dt} = b (X - I) \qquad (4.34)$$

where $X = \frac{Va \sigma_e}{l}$, $b = \frac{Va \sigma_f}{Nq l}$

Integrating equation (4.34),

$$I = X - Ye^{-bt} \qquad (4.35)$$

where Y is a constant of integration.

Equation (4.35) is identical with the empirical equation (4.17) obtained during the erosion of an aluminium oxide film.

The impedance of a highly insulating film may be obtained in a similar manner. In the cell in which the impedance was measured the film was eroded at constant current rather than constant voltage. For this case equation (4.29) becomes,

$$I = \frac{V(t) a}{N l} \left\{ \sigma_e (N - n) + \sigma_f n \right\} \qquad (4.36)$$

and equation (4.31),

$$\frac{dn}{dt} = - \frac{V(t)\, a\, \sigma_f\, n}{N q\, l} \qquad (4.37)$$

Combining equations (4.36) and (4.37),

$$\frac{dn}{dt} = - \frac{\sigma_f n}{q} \left\{ \sigma_e (N-n) + \sigma_f n \right\} \qquad (4.38)$$

Now, $\sigma_e \gg \sigma_n$ so equation (4.38) may be written approximately,

$$\frac{dn}{dt} = - \frac{\sigma_e \sigma_f n (N-n)}{q} \qquad (4.39)$$

or $$\int \frac{dn}{n} + \int \frac{dn}{N-n} = - \frac{N \sigma_e \sigma_f}{q} \int dt . \qquad (4.40)$$

Integrating equation (4.40),

$$\log_e \left(\frac{k_2 n}{N-n} \right) = - b_1 t \qquad (4.41)$$

where $b_1 = \frac{N \sigma_e \sigma_f}{q}$; k_2 = a constant of integration.

As before, assuming that $\sigma_e \gg \sigma_f$, the resistance of the film when n of the elementary parts are covered by film,

$$R(n) = \frac{Nl}{\sigma_e a (N-n)} \qquad (4.42)$$

Equation (4.41) may be written,

$$n = \frac{N e^{-b_1 t}}{k_2 + e^{-b_1 t}} \qquad (4.43)$$

Hence, from equation (4.42) and (4.43),

$$R(t) = \frac{1}{\sigma_e q} \left(k_2 + e^{-b_1 t} \right)$$

If the current is applied in pulses of duration t_1 , then $t = t_1 \bar{N}$

Where $\bar{N}$ is the number of pulses.

Hence, $R = X_1 + Y_1 e^{-b_2 \bar{N}}$. (4.44)

where $X_1 = \frac{k_2 l}{\sigma_e a}$; $Y_1 = \frac{l}{\sigma_e a}$; $b_2 = \frac{N \sigma_e \sigma_f t_1}{q}$

Equation (4.44) is identical with the expression (4.19).

The capacitance due to n elementary parts of film will be, (l' is the final thickness of the film, $l' \ll l$),

$$C = \frac{\epsilon a}{4\pi N} \left\{ \frac{N}{l'} - \left(\frac{1}{l'} - \frac{1}{l}\right) n \right\} \quad (4.45).$$

Substituting for n from equation (4.41),

$$C = \frac{\epsilon a}{4\pi} \left\{ \frac{1}{l'} - \left(\frac{1}{l'} - \frac{1}{l}\right) \frac{e^{-b_1 t}}{k_2 + e^{-b_1 t}} \right\} \quad (4.46)$$

Substituting for b_1 and t, and writing $k_1 = \frac{\epsilon a b_2}{4\pi} \left(\frac{1}{l'} - \frac{1}{l}\right)$

and $C_0 = \frac{\epsilon a}{4\pi} \left\{ \frac{k_2 + 1}{l} - \frac{k_2}{l'} \right\} \quad (4.47)$

If $b_2 \ll 1$ this becomes for small $\bar{N}$,

$$C = C_0 + k_1 \bar{N} \quad (4.48).$$

This again is the same as expression (4.20).

It is not expected that this approximate form should hold if a very large number of pulses were employed.

4 (13) Theories of rectification.

A number of theories purporting to explain the different conductivities in the forward and reverse directions of aluminium oxide films have been advanced. Scholte and van Geel (1953) for example, consider the oxide film to consist of three regions; an n -type semi-conductor next to the aluminium, a central uniform region, and a p-type semi-conductor next to the electrolyte. Applying a voltage across the film in the direction which makes the aluminium negative causes the electrons and holes to drift into the

central region. As recombination does not occur instantly the number of charge carriers, and therefore the conductivity, increases.

Vermilyea (1956) has observed streams of hydrogen gas bubbles given off from the surface of an oxide film. These spots were often situated at inclusions in the metal. He concludes that the conduction takes place mainly through defects in the film.

It is possible to explain the rectifying action of anodic oxide films if the mechanism postulated in section (4.12) is assumed. If this is true it is not necessary to assume that the conductivity of the film is different in different directions. There is some indication that this is so. Berry and Sloan (1959) have developed some dry, tantalum capacitors by forming films on tantalum anodes electrolytically and then allowing them to dry. The other plate of the capacitor was made by evaporating metal on to the exposed surface of the oxide film. It was found that the leakage resistance of these capacitors did not show an abrupt change when the potential of the tantulum plate was taken from positive to negative.

In order to discover if this observation of Berry and Sloan was also true of aluminium oxide films the data shown in Figure (4.18) were obtained. Films were formed on aluminium to various voltages V_f, then removed from

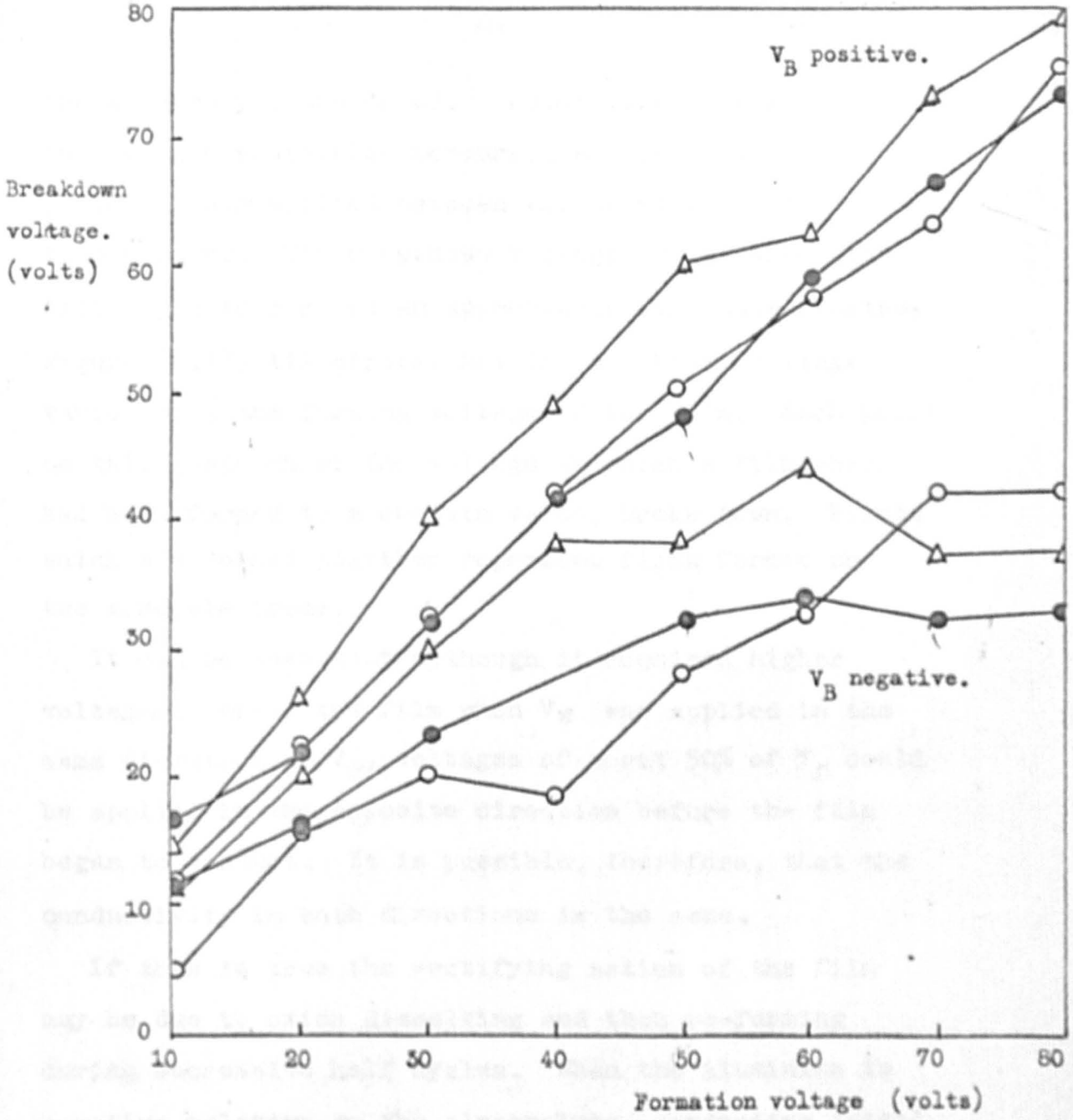

80
70
Breakdown
voltage.
(volts)
60
50
40
30
20
10
0
V_B positive.
V_B negative.
10
20
30
40
50
60
70
80
Formation voltage (volts)

Figure 4.17.

the electrolyte and dried. Later they were inserted into a tube containing mercury. A slowly increasing potential was applied between the aluminium metal and this mercury. The breakdown voltage V_B at which the film began to conduct an appreciable current was noted. Figure (4.17) illustrates how the breakdown voltage varied with the forming voltage of the film. Each point on this graph shows the voltage at which a film, which had been formed to a certain value, broke down. Points which are joined together represent films formed on the same electrode.

It can be seen that although it required higher voltages to break the film when V_B was applied in the same direction as V_f, voltages of about 50% of V_f could be applied in the opposite direction before the film began to conduct. It is possible, therefore, that the conductivity in both directions is the same.

If this is true the rectifying action of the film may be due to oxide dissolving and then re-forming during successive half cycles. When the aluminium is negative relative to the electrolyte, conducting 'pits' which pass a large current, are rapidly formed. As the electrode becomes positive the ions are channelled into these pits and quickly fill them with highly insulating oxide. The film then presents a uniform high resistance to the current for the rest of this half cycle.

4 (14) Discussion of results.

The purpose of this chapter has been to give an account of the properties of anodic oxide films. In particular it has been found that the impedance of these films can be changed by passing electric currents through them. The way in which this impedance varies with the charge passed has been calculated for the case in which the film dissolves electrochemically. The A.C. resistance of these films was shown to be (Equation 4.44)

$$R = X_1 + Y_1 \exp(-k_2 \bar{N}),$$

and the capacitance (Equation 4.48)

$$C = C_0 + k_1 \bar{N},$$

where $\bar{N}$ is the number of constant current pulses passed through the film.

The current through a film being eroded at constant applied voltage was shown to be (Equation 4.35)

$$I = X - Y \exp(-bt).$$

It is still possible, however, that instead of dissolving, the film may be broken down by bubbles of hydrogen, formed on the metal / oxide interface, cracking the film and allowing the electrolyte to come into contact with the electrode metal. This would be more in accord with the observation of Vermilyea (1956). As the quantity of hydrogen produced would be determined by the charge passed according to Faraday's law, and as the change in conductance of each part of the film would be a step function, the results derived in section (4.12) would still be applicable.

REFERENCES, CHAPTERIV.

Berry, R.W. and Sloan, D.J.(1959) "Tantalum Printed Capacitors". Proc. I.R.E., vol. 47, no.6. pp. 1,070 - 75.

Cabrera, N and Mott, N.F. (1948) Repts. Prog. Physics, vol. 12, p. 163.

Charlesby, A (1953a) "Ionic Currents and Film Growth in Thin Oxide Layers on Aluminium." Proc. Phys. Soc. B, vol. 66, pp 317 - 329.

Charlesby, A. (1953b) "Electron Currents in Thin Oxide Films on Aluminium." Proc. Phys. Soc. B, vol. 66, pp 533 - 541.

Dewald, J.F. (1954), Acta Met., vol. 2, p. 340.

Dewald, J.F. (1955), J. Electrochem. Soc., vol. 102. p. 1.

Geel, W. Ch. van and Pistorius, C.A. (1956) "On the Residual Voltage with Electrolytic Capacitors." Phillips Res. Rep., vol. 11, pp 471 - 478.

Guntherschulze, A and Betz, H. (1937) "Elektrolytkondensatoren." Krayn, Berlin.

Hoar, T.P. (1959) "Anodic Behaviour of Metals", in "Modern Aspects of Electrochemistry, No. 2 " Bockris (Ed.) Butterworths, London, pp 262 - 342.

MacKay D.M. and Ainsworth, W.A. (1961) "Improvements in or relating to Automatic Controlling Devices and the Like." Brit. Prov. Pat. Spec. 01288/61.

MacKay, D.M. (1961) "Information and Learning". in "Lernende Automaten", Oldenburg, Munich, pp. 40 - 49.

Mott, N.F. (1947) "Theory of the Formation of Protective Oxide Films on Metals." Trans. Faraday Soc., vol. 43, pp 429 - 434.

REFERENCES, CHAPTER IV (Continued)

Scholte, J.W.A. and Geel, W. Ch. van (1953) "Impedance of the Electrolytic Rectifier." Phillips Res. Rep., vol 8, pp 47 - 72.

Smith, A.W. (1959) "The Impedance, Rectification and Electroluminescence of Anodic Oxide Films on Aluminium." Canad. J. Phys., vol. 37, pp 591 - 606.

Vermilyea, D.A. (1956) "On the Mechanism of Electrolytic Rectification." J. of App. Phys., vol 27, no. 8, pp 963 - 4.

Verwey, E.J.W. (1935) Physica, vol. 2, p. 1,059.

Young, L. (1954) "Anodic Oxide Films, Parts I, II and III." Trans. Faraday Soc., vol. 50, pp 153 - 171.

Young, L, (1955) "Anodic Oxide Films, Part IV". Trans. Faraday Soc., vol. 51, pp 1250 - 60.

Young, L. (1956) "Anodic Oxide Films, Part V". Trans. Faraday Soc., vol. 52, pp 502 - 14.

Young, L. (1961) "Anodic Oxide Films" Academic Press, London.

CHAPTER V

DEVICES EMPLOYING ANODIC OXIDE FILMS

5 (1) Introduction

Variable impedance cells in which the impedance is controlled by the growth of insulation are formally equivalent to those in which the impedance is controlled by the growth of conducting pathways. There are, however, certain applications for which the special properties of these cells make them more suitable.

From a general theory of self-adaptive organisation (MacKay, 1955, 1956) it seems desirable for a system to begin to explore its universe with relatively few degrees of freedom, and to increase the diversity of its repertoire of actions as it builds up its knowledge. One of the simplest ways in which this behaviour could be produced might be to couple the circuits controlling the 'effectors' with electrolytic cells in which insulation could be grown as the system developed. Anodic oxidation processes lend themselves to such applications since the insulating film is easier to grow than dissolve. Alternating currents can be used for signalling and direct currents for permanently isolating a particular control system. It might also be arranged that circuits which were most often used by themselves became independent automatically as the rectifying action

of the cells tends to promote the growth of films when alternating currents are used.

The compactness of the film is advantageous when another principle is employed. It has been suggested by MacKay (1961) that whole adaptive networks may be switched in or out of circuits by selectively replacing or draining the electrolyte from electrolytic cells, provided that cells whose impedance is controlled by an insulating film regain their original impedance when the electrolyte is restored.

This chapter starts with a discussion of the factors which affect the impedance-changing properties of cells containing aluminium oxide films. The stability of the impedance is then considered, and the remainder of the chapter is concerned with the construction and properties of circuits containing large numbers of cells of this type.

5 (2) Parameters affecting the "rate of growth" of oxide films

A measure of the rate of growth of oxide film on aluminium may be obtained by observing the current through the cell as a function of time with constant applied voltage. Equation (4.14) may be written approximately,

$$I\ (t + t_o) = D \qquad (5.1)$$

for small t. (On the occasions in which this is used in this section I_o, which is neglected, is only about 3% of I).

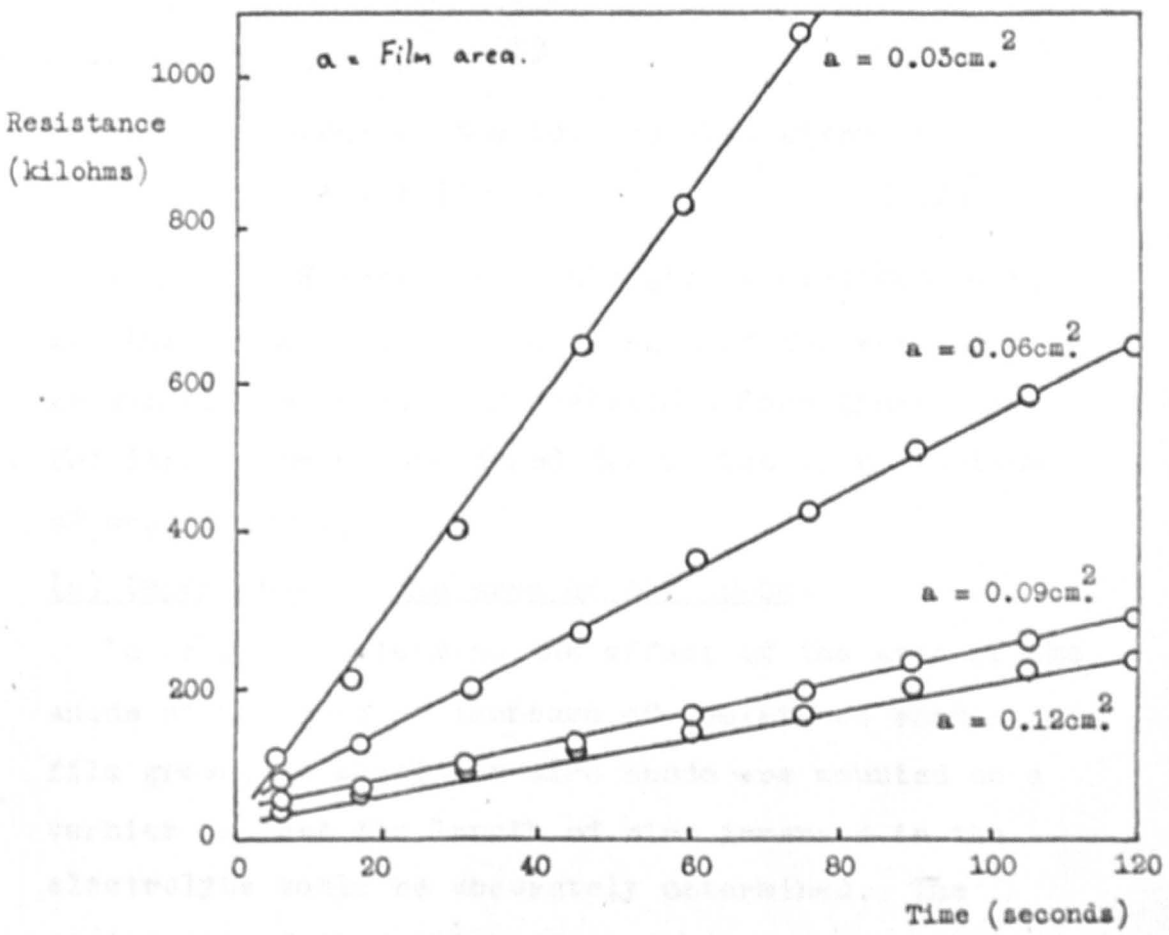


Figure 5.1.

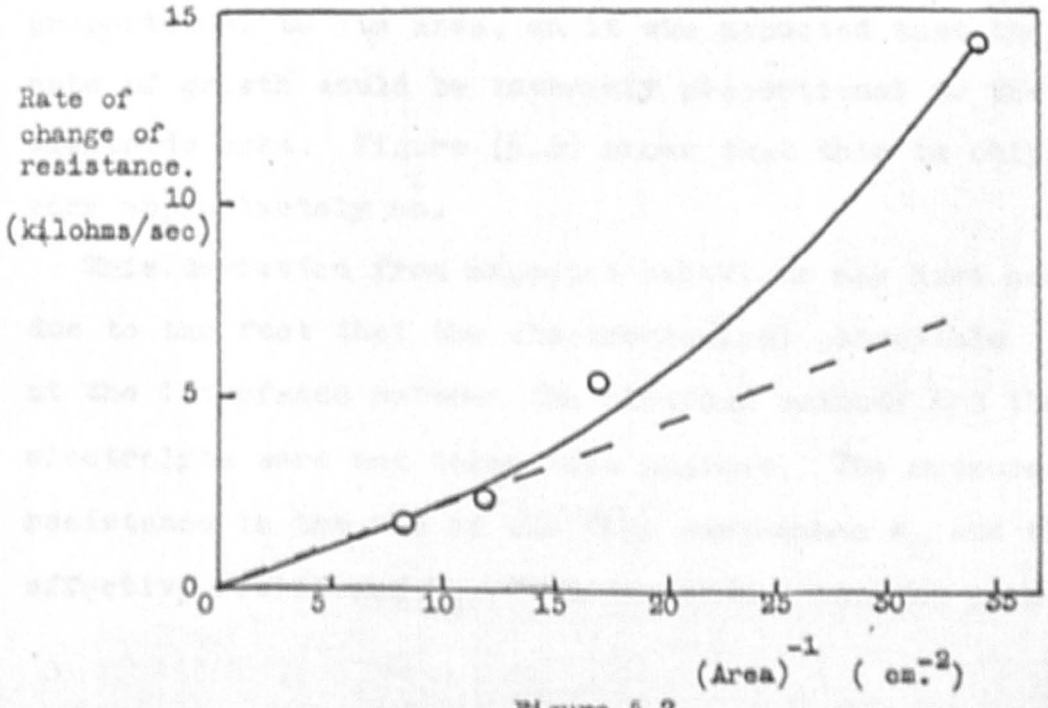


Figure 5.2.

The resistance of the cell is then given by,

$$R = \frac{V}{D} \quad (t + t_o) \qquad (5.2)$$

A plot of R versus t should give a straight line, and the slope of this is a measure of the rate of growth of the film. Any deviations from linearity for large t were considered due to the approximation of neglecting I_o.

(a) Dependence on the area of the anode.

In order to determine the effect of the area of the anode on the rate of increase of resistance when a film grows, an aluminium wire anode was mounted on a vernier so that the length of wire immersed in the electrolyte could be accurately determined. The resistance of the cell was measured as a function of time, and the results plotted as shown in Figure (5.1)

The resistance of a film is normally inversely proportional to its area, so it was expected that the rate of growth would be inversely proportional to the electrode area. Figure (5.2) shows that this is only very approximately so.

This deviation from expected behaviour may have been due to the fact that the electrochemical potentials at the interfaces between the platinum cathode and the electrolyte were not taken into account. The measured resistance is the sum of the film resistance R_f and the effective resistance R_e of the remaining current path

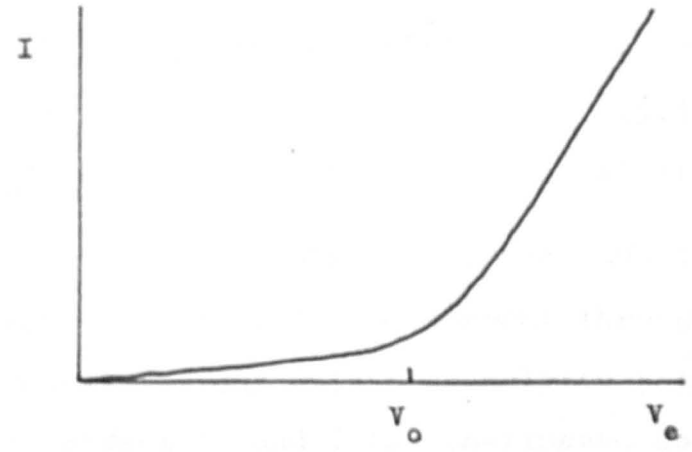


Figure 5.3.

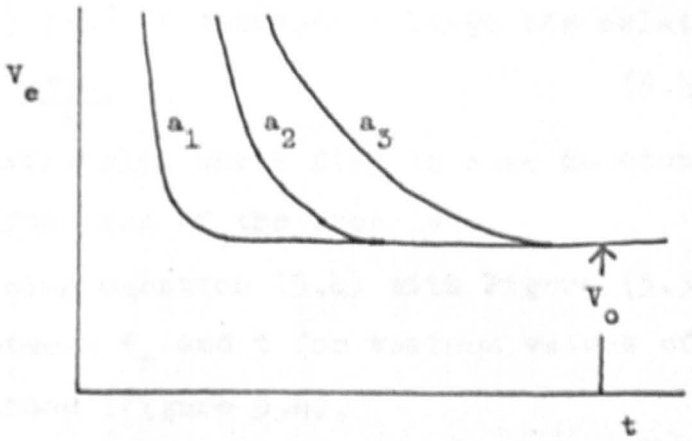


Figure 5.4.

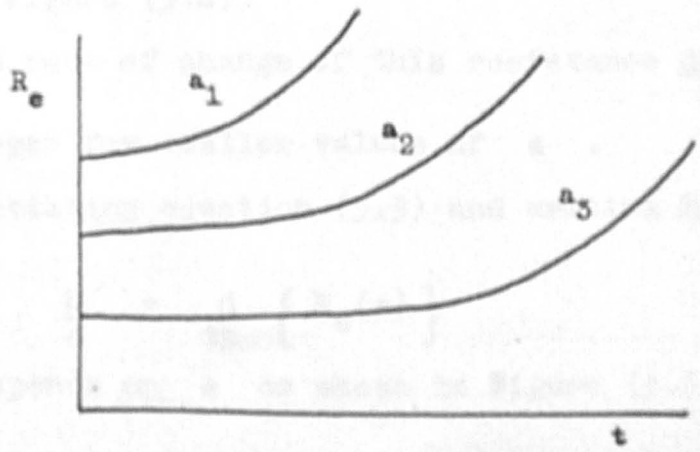


Figure 5.5.

through the electrolyte to the platinum electrode,

$$R = R_f + R_e \qquad (5.3)$$

Now $R_e = \frac{Ve}{I}$ where V_e is the potential difference between the oxide / electrolyte interface and the platinum electrode, and I is the current through the cell. In general for any such electrolytic path the relationship between V_e and I is non-linear, having roughly the form shown in Figure (5.3).

It can be seen from the experimental results (Figure 5.1) that at constant voltage the relation

$$I = \frac{f(a)}{t} \qquad (5.4)$$

holds approximately, where f(a) is some monotonically increasing function of the area a .

By combining equation (5.4) with Figure (5.3) a relation between V_e and t for various values of a can be obtained (Figure 5.4).

R_e as a function of time is then seen to be as in Figure (5.5) which was produced by combining equation (5.4) with Figure (5.4).

The mean rate of change of this resistance $\frac{dRe}{dt}$ will thus be larger for smaller values of a .

Differentiating equation (5.3) and writing $R_f = \frac{K'(t+t_0)}{a}$

$$\frac{dR}{dt} = \frac{K'}{a} + \frac{d}{dt}\left\{R_e(a)\right\}$$

where R_e depends on a as shown in Figure (5.5)

The measured "rate of growth" may be expected to be proportional to the electrode area only for relatively large areas.

(b) Dependence on the concentration of the electrolyte.

The rate at which a reaction goes to completion is limited by the slowest, or rate-controlling, step. In the growing of an oxide film the rate of growth may be determined by the transport of aluminium ions through the film, the conduction of oxygen ions through the electrolyte, or by a chemical reaction at one of the interfaces. The concentration of the electrolyte should to some extent determine the supply of oxygen ions at the film surface.

In order to find out if the concentration of ammonium borate in the electrolyte had any effect on the growth of the film, the growth rate was measured for a number of cells containing different concentrations of solute. The maximum concentration was 133 gm. per litre, at which concentration the solution was saturated, and the minimum was 0.5 gm. per litre, which seemed to be the lowest concentration at which a film was formed. The results are shown in Figure (5.6) which is a plot of the resistance of a cell versus time for various concentrations in this range.

Figure (5.7) shows how the rate of resistance change varies with the concentration of ammonium borate in the

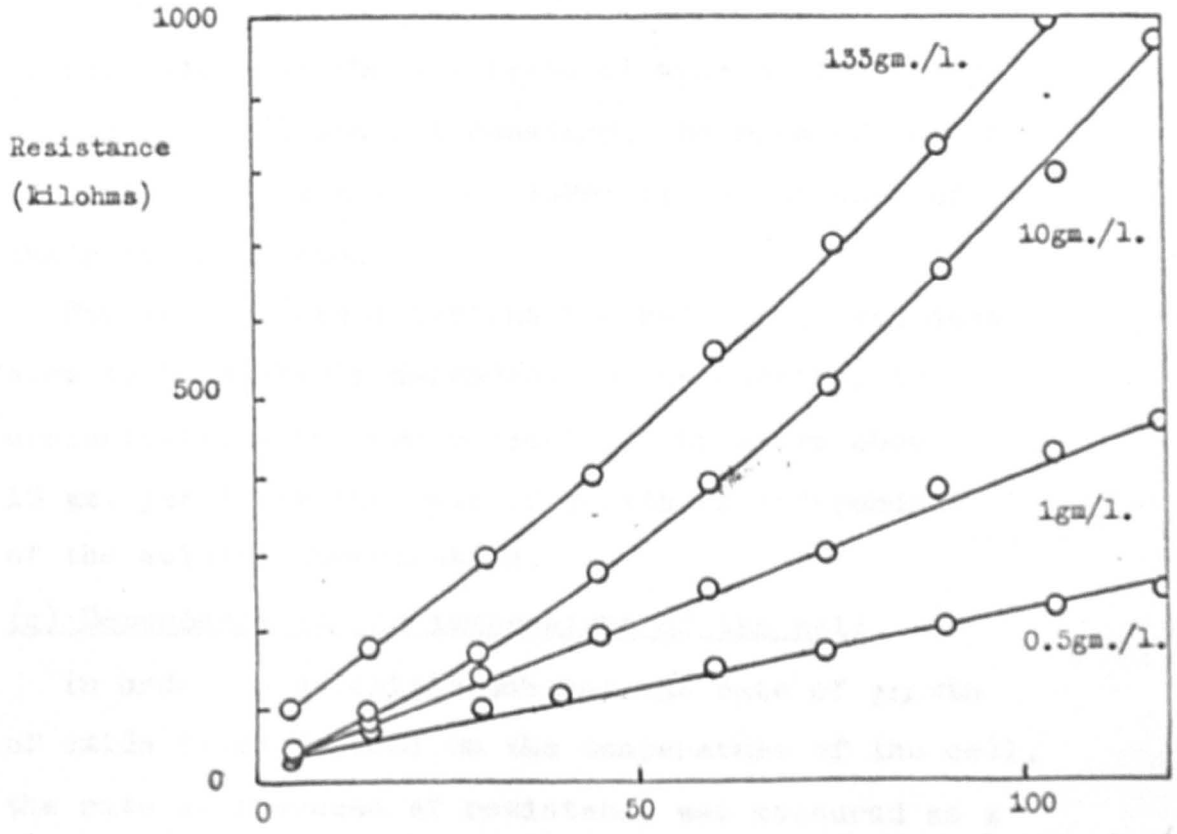


Figure 5.6.

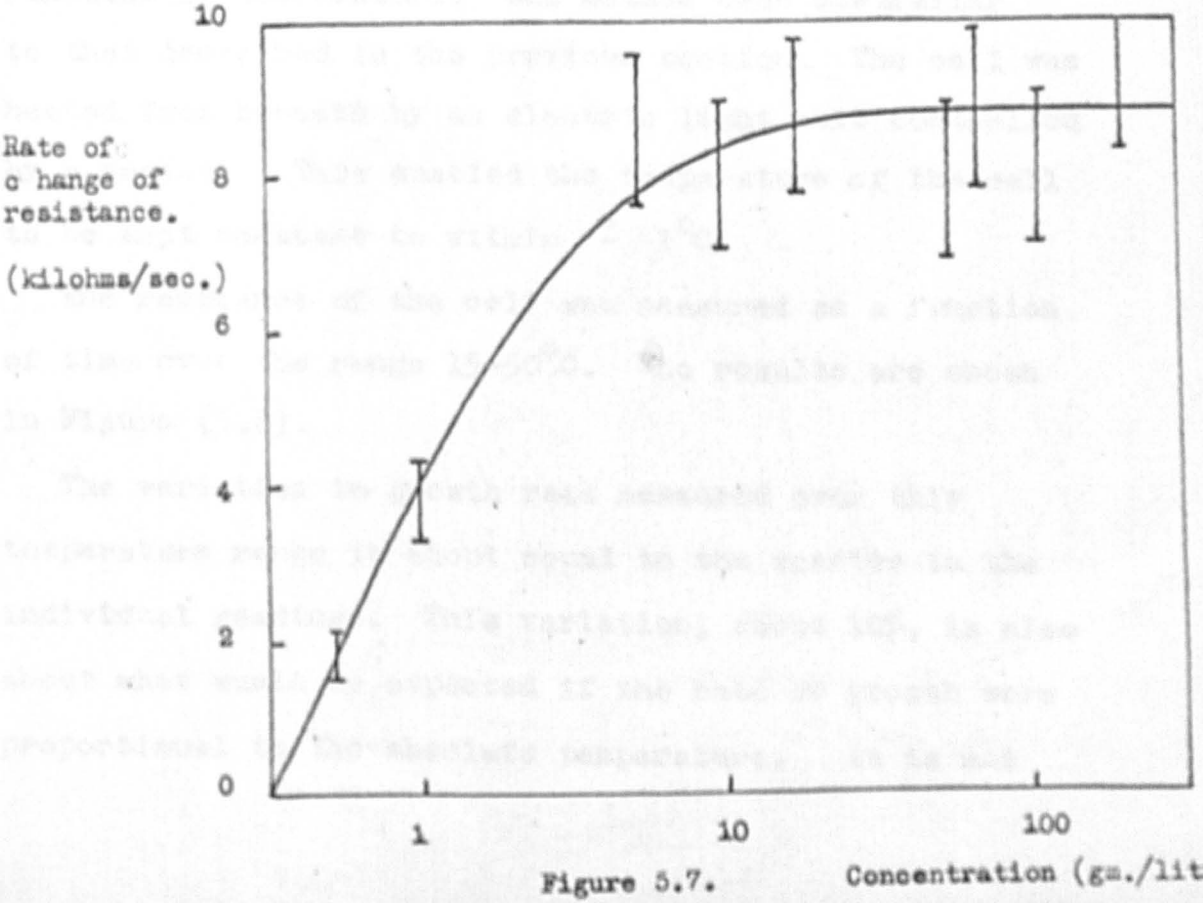


Figure 5.7.

electrolyte. As the gradients of some of the lines in Figure (5.6) are not constant, the rate of change of resistance at t = 0 was taken as the measure of the rate of growth.

For very dilute solutions the rate of growth does seem to be markedly dependent on the electrolyte concentration, but for concentrations above about 10 gm. per litre the rate of growth is independent of the solute concentration.

(c) Dependence on the temperature of the cell.

In order to ascertain whether the rate of growth of oxide films depends on the temperature of the cell, the rate of increase of resistance was measured as a function of temperature. The method used was similar to that described in the previous section. The cell was heated from beneath by an electric light bulb controlled by a variac. This enabled the temperature of the cell to be kept constant to within $\pm$ 1°C.

The resistance of the cell was measured as a function of time over the range 15-50°C. The results are shown in Figure (5.8).

The variation in growth rate measured over this temperature range is about equal to the scatter in the individual readings. This variation, about 10%, is also about what would be expected if the rate of growth were proportional to the absolute temperature. It is not

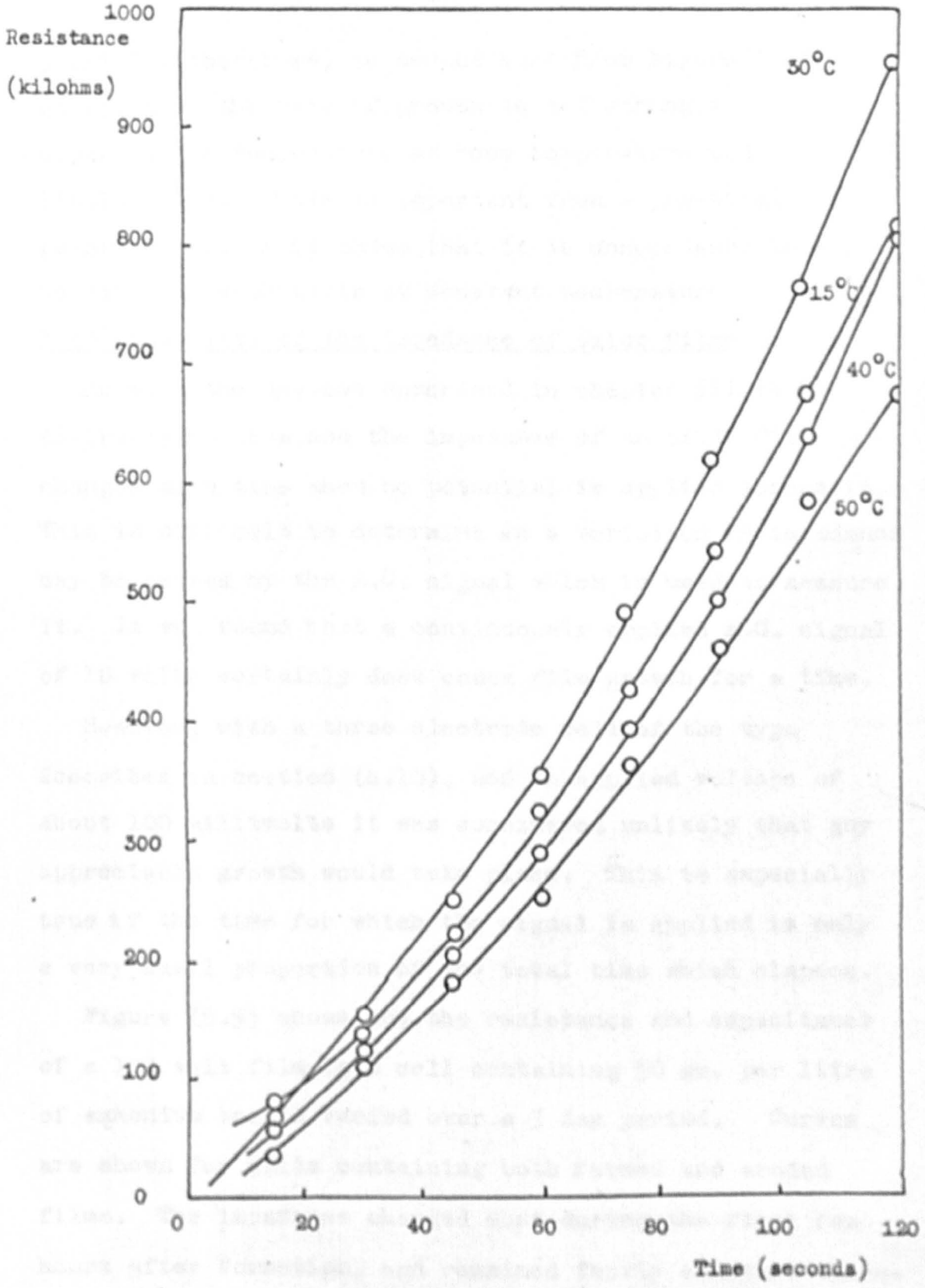

Resistance
(kilohms)
1000
900
800
700
600
500
400
300
200
100
0
0
20
40
60
80
100
120
Time (seconds)
30°C
15°C
40°C
50°C

Figure 5.8.

possible, therefore, to deduct much from Figure (5.8) except that the rate of growth is not strongly dependent on temperature at room temperature and a little above. This is important from a practical point of view as it shows that it is unnecessary to try to maintain such cells at constant temperature.

5 (3) Stability of the impedance of oxide films.

As with the devices described in chapter III it is desirable to know how the impedance of an oxide film changes with time when no potential is applied across it. This is difficult to determine as a variation in impedance may be caused by the A.C. signal which is used to measure it. It was found that a continuously applied A.C. signal of 10 volts certainly does cause film growth for a time.

However, with a three electrode cell of the type described in section (4.10), and an applied voltage of about 100 millivolts it was considered unlikely that any appreciable growth would take place. This is especially true if the time for which the signal is applied is only a very small proportion of the total time which elapses.

Figure (5.9) shows how the resistance and capacitance of a 100 volt film in a cell containing 50 gm. per litre of ammonium borate varied over a 3 day period. Curves are shown for cells containing both formed and eroded films. The impedance changed most during the first few hours after formation, and remained fairly constant thereafter.

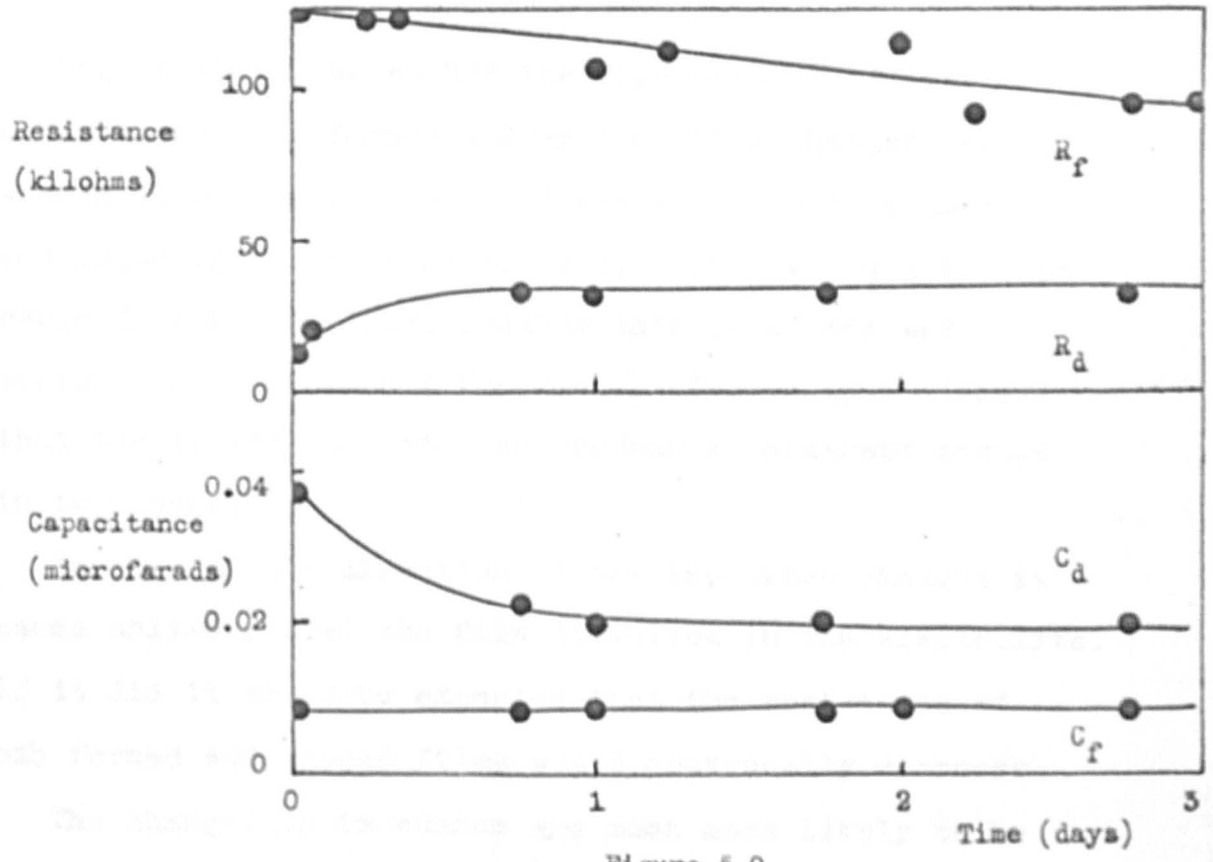


Figure 5.9.

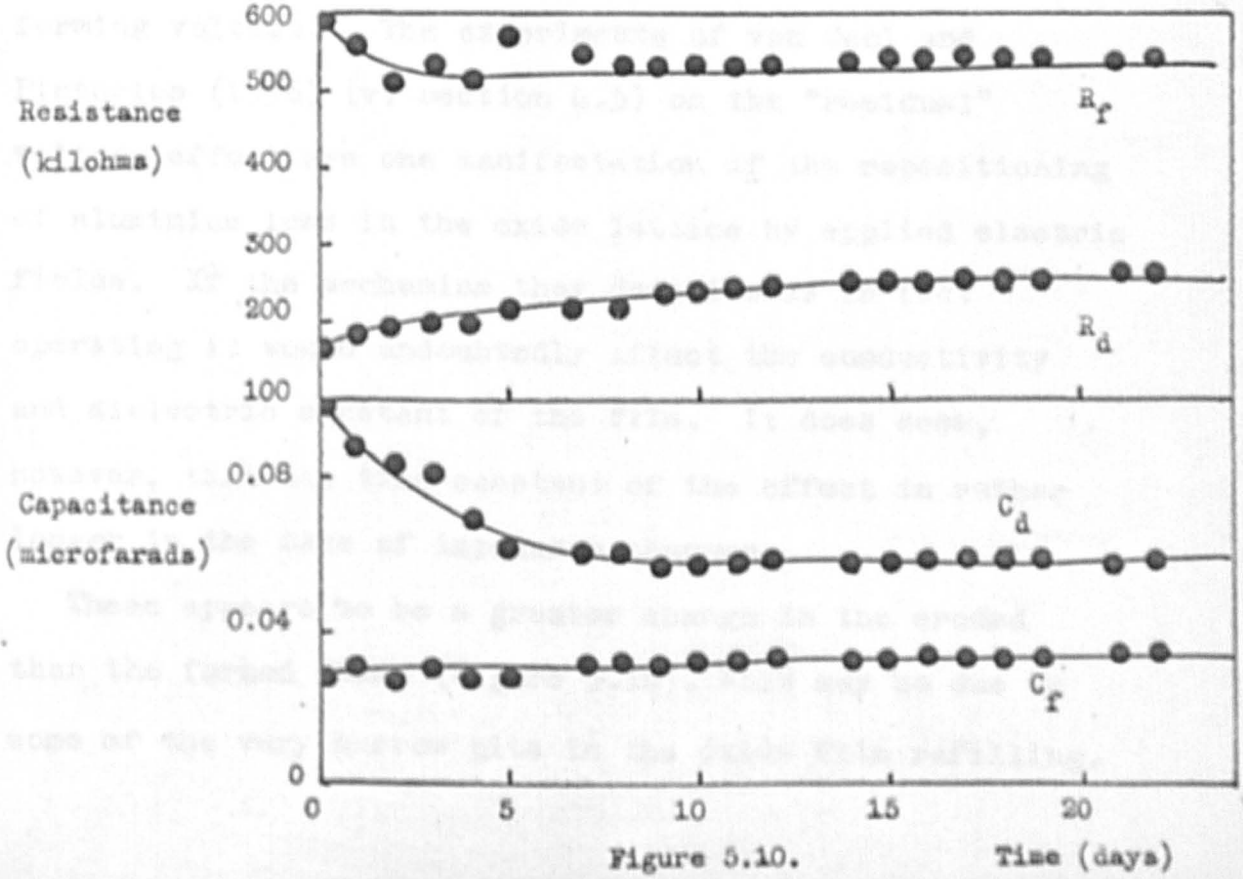


Figure 5.10.

Figure (5.10) shows how the impedance of similar cells containing formed and eroded films changed over a much longer period (about 3 weeks). The impedance and capacitance of these cells is shown over a different range from Figure (5.9) because this impedance was measured at a different frequency. Figure (5.10) shows that the erosion current does produce a permanent change in impedance.

In view of the direction of the impedance changes it seems unlikely that the film dissolves in the electrolyte. If it did it would be expected that the resistance of both formed and eroded films would continually decrease.

The changes in impedance are much more likely to be due to ionic space charge in the film set up by the forming voltage. The experiments of van Geel and Pistorius (1956) (v. section 4.5) on the "residual" voltage effect are one manifestation of the repositioning of aluminium ions in the oxide lattice by applied electric fields. If the mechanism they describe is in fact operating it would undoubtedly affect the conductivity and dielectric constant of the film. It does seem, however, that the time constant of the effect is rather longer in the case of impedance changes.

There appears to be a greater change in the eroded than the formed films (Figure 5.10). This may be due to some of the very narrow pits in the oxide film refilling.

As there is no current flowing to remove the insulating oxide this would lead to an increase in the film resistance and a decrease in capacitance.

5 (4) Effect of a copper electrode.

In order that the mechanism of the formation of oxide films on the anode of a cell could be elucidated an insoluble counter-electrode was employed at first. A platinum electrode was used as this fulfilled this condition ideally, but carbon appeared to be equally good. However, it is possible to produce matrices of variable impedances by weaving together aluminium and cotton-covered copper wires, with electrolyte absorbed in the cotton covering. (MacKay and Ainsworth, 1961). These devices enable full use to be made of the parallel switching principle (Section 5.1). The possibility of replacing the platinum electrode with one of copper was therefore investigated.

One of the requirements of a device used in a system containing many hundreds of adaptive elements is that it should be inexpensive. This provided a further reason for studying cells in which the counter-electrodes were of copper.

When the platinum electrode of a simple cell was replaced by a copper one it was found that the oxide film could be grown on the aluminium as before. But, when the current through the cell was reversed, an oxide of copper

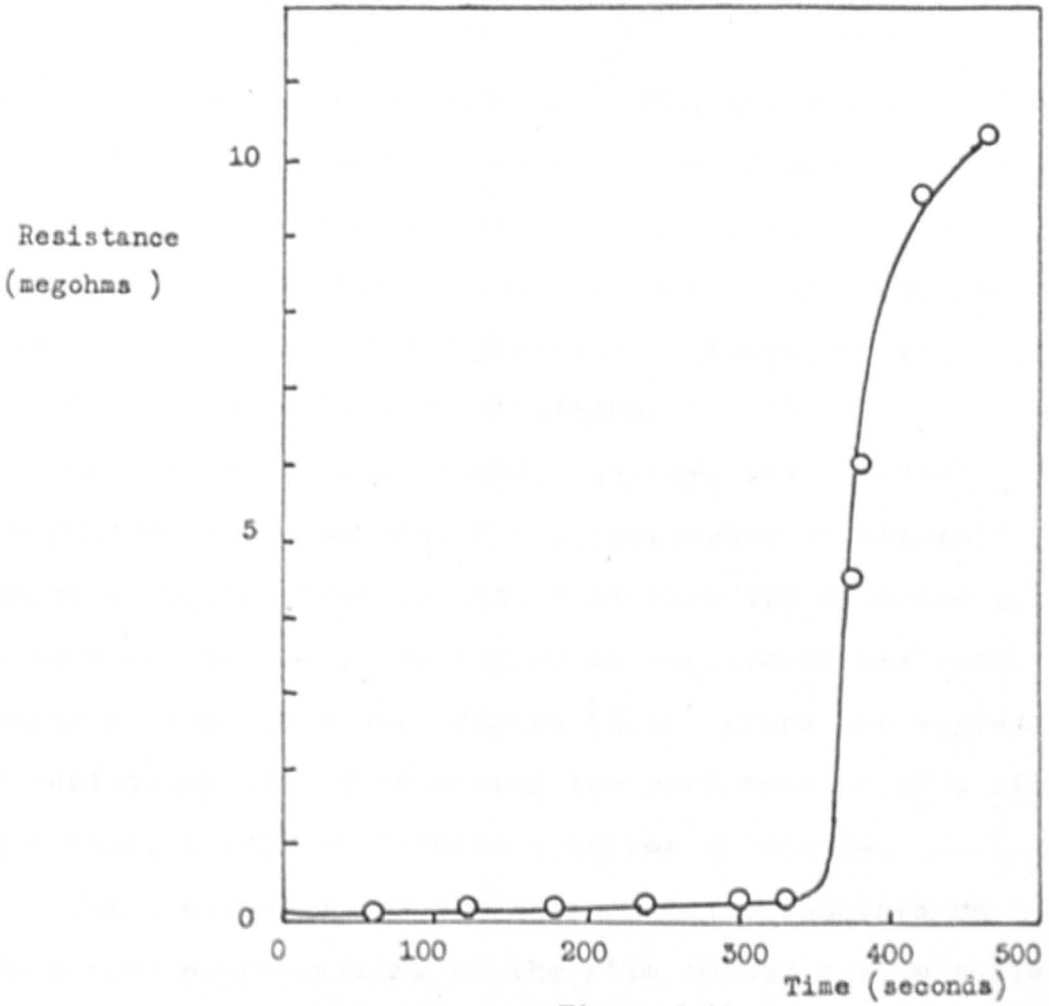


Figure 5.11.

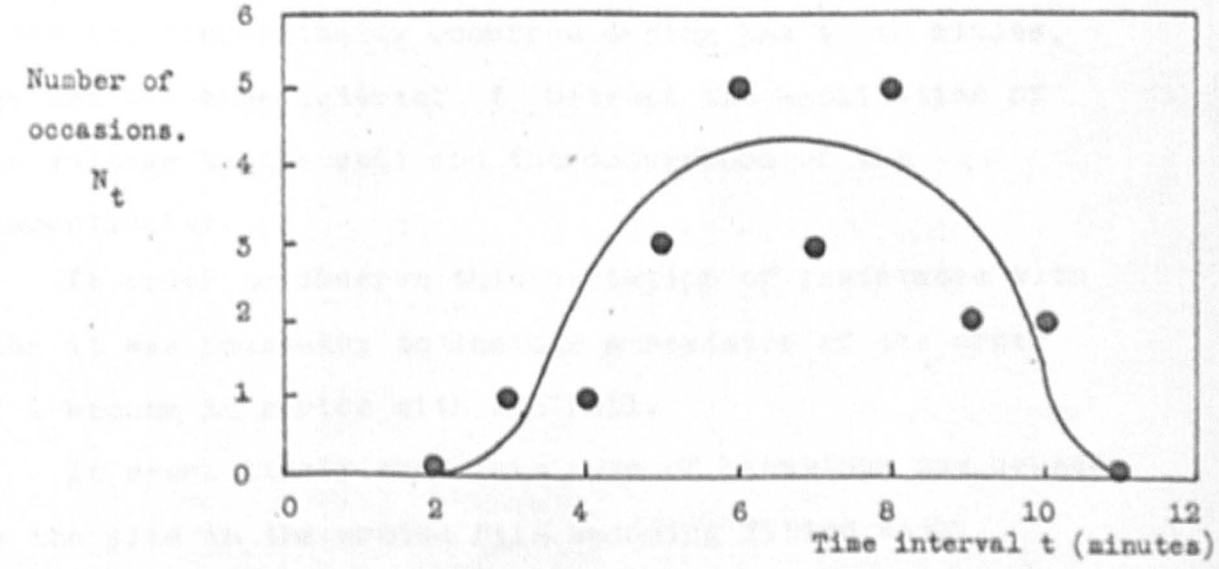


Figure 5.12.

formed on the copper electrode. This eventually isolated the copper electrode from the electrolyte and raised the resistance of the cell again. It was found that the addition of ammonia to the electrolyte caused this oxide to be dissolved. This allowed the cell to function more like the one with the platinum electrode.

One important difference, however, was observed. It was noticed that instead of the resistance of the cell changing fairly smoothly with time when the film was grown at constant voltage, the change in resistance was more nearly a step function. Figure (5.11) shows the variation of resistance with time during the re-formation of a film in a typical cell containing a copper electrode.

The times at which the discontinuity occurred on successive re-formations of the film showed a wide variation. Figure (5.12) is a plot of the number of occasions N_t on which the discontinuity occurred during the t th minute, against the time interval t between the application of the voltage to the cell and the occurrence of the discontinuity.

In order to observe this variation of resistance with time it was necessary to include a resistor of the order of 1 megohm in series with the cell.

It seems likely that this type of behaviour was caused by the pits in the eroded film becoming filled with metallic copper deposited during the break down of the film.

During this process the copper electrode was the anode of the cell and it was noticed that the aluminium electrode became copper coloured. This suggests that copper was dissolved from the anode and deposited on the cathode. When the aluminium electrode was made positive the deposited copper redissolved. The oxide film could not form in a pit until all the copper had gone. The series resistor limited the current so that it took several minutes for enough charge to pass for all the copper to redissolve. The conductivity of the copper is so much higher than that of the oxide that the resistance of the cell would not change much until the last pit was being refilled with oxide. The actual filling would then take place very quickly, leading to the sharp rise in resistance observed.

5 (5) Drying characteristics of cells.

Because of these complications introduced by the use of copper electrodes the investigation of the woven mesh of aluminium and copper wires was discontinued. This mode of construction is suitable, however, for manufacturing matrices of cells in which only uni-directional changes in impedance are required. Further investigations were directed towards producing matrices of cells in which the possibilities of bi-directional impedance changes and the parallel switching principle could be exploited.

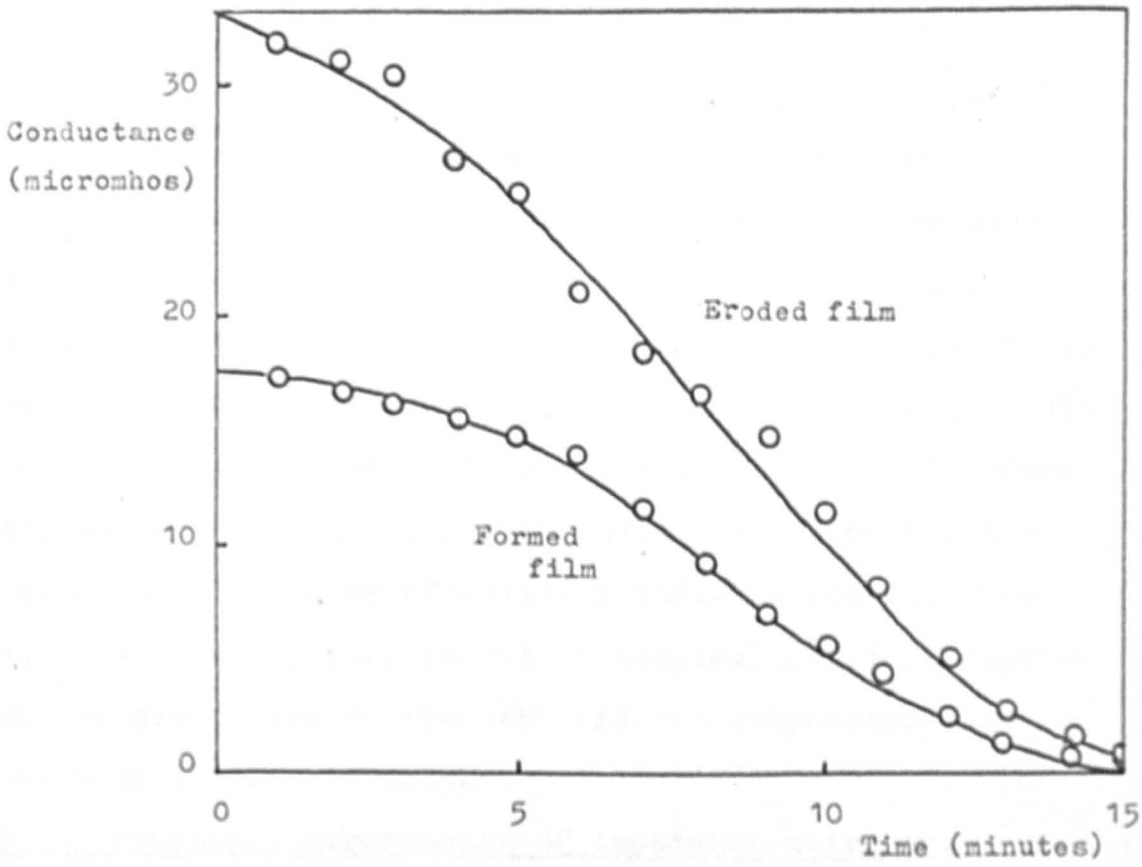


Figure 5.13.

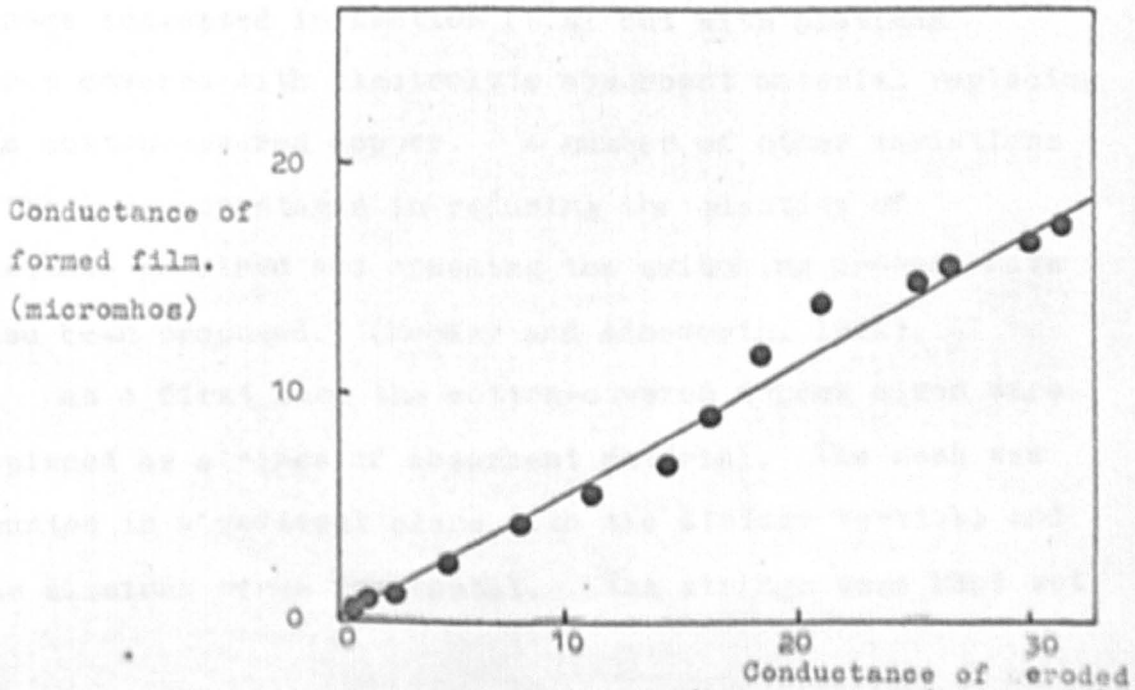


Figure 5.14.

A cell was constructed consisting of electrodes of aluminium and platinum separated by a piece of filter paper moistened by electrolyte. This was used to study the drying characteristics of films. Figure (5.13) shows how the D.C. conductance of formed and eroded films varied during the drying process, and Figure (5.14) shows that the ratio of the conductances of the films in these two states remained fairly constant. The time for the resistance to become effectively infinite (conductance zero) was rather long (about 15 minutes) but this depends on the dimensions of the cell and the temperature and humidity of the atmosphere.

5 (6) Practical embodiments of impedance matrices.

(a) Strings.

Adaptive matrices in which the impedances may be increased or decreased may obviously be made using the scheme suggested in section (5.4) but with platinum wires covered with electrolyte absorbent material replacing the cotton-covered copper. A number of other variations which have advantages in reducing the quantity of platinum required and speeding the switching process have also been proposed. (MacKay and Ainsworth, 1961).

As a first step the cotton-covered copper wires were replaced by strings of absorbent material. The mesh was mounted in a vertical plane with the strings vertical and the aluminum wires horizontal. The strings were kept wet

with a trickle of electrolyte from a reservoir situated above, and electrical contact was made to each string with a small, platinum electrode. This arrangement worked tolerably well, but the impedance of the strings was rather high. It is obviously desirable that the resistance of the wet string should be less than the lowest value of the resistance of a junction, between the string and an aluminium wire. The resistance of the string may be decreased by increasing the thickness of the string and the rate of flow and conductivity of the electrolyte. It was found empirically that the conductance of a string σ depended upon the rate of flow F, according to,

$$\sigma = \sigma_0 + c_0 F^{\frac{1}{2}} \qquad (5.6)$$

where σ_0 and c_0 are constant for a given string. This relation, however, breaks down if the electrolyte is actually flowing, not dripping.

The network could be made electrically inactive simply by turning off the supply of electrolyte. The time required for the strings to dry out was about 5 minutes.

Two types of absorbent material were tested as the vertical strings - cotton and wool. It was found that cotton dried more quickly than wool, but that wool had more stable electrical properties. It absorbed the drops of liquid so its resistance did not show periodic

fluctuations as did that of the cotton strings.

(b) Tubes.

A more satisfactory arrangement was produced by replacing the strings by insulated plastic tubes containing the electrolyte. These were pierced transversely by the aluminium wires as shown in Figure (5.15). This arrangement has the advantage that the liquid is totally enclosed, and it can be made inactive or active as rapidly as the tubes can be emptied, or refilled. It was found that the impedance of the junctions returned to within 10% of their initial values when the electrolyte was restored.

There are a number of ways in which the electrolyte may be brought to and from these cells. (MacKay and Ainsworth, 1961). One method which was tried was to connect a compressible bulb filled with electrolyte to the lower end of each tube. These bulbs can be compressed electro-mechanically, causing the electrolyte to flood the cell and the conductance of the junctions to become active. Another possibility is to confine the transverse aluminium wires to one region of the cell, and to tilt or invert the tube in order to control the distribution of electrolyte.

Tubes which were wide enough to make the impedance of the column of the liquid small compared with those of the junctions were used in these experiments. It was then found possible to obtain nearly the same range in impedance

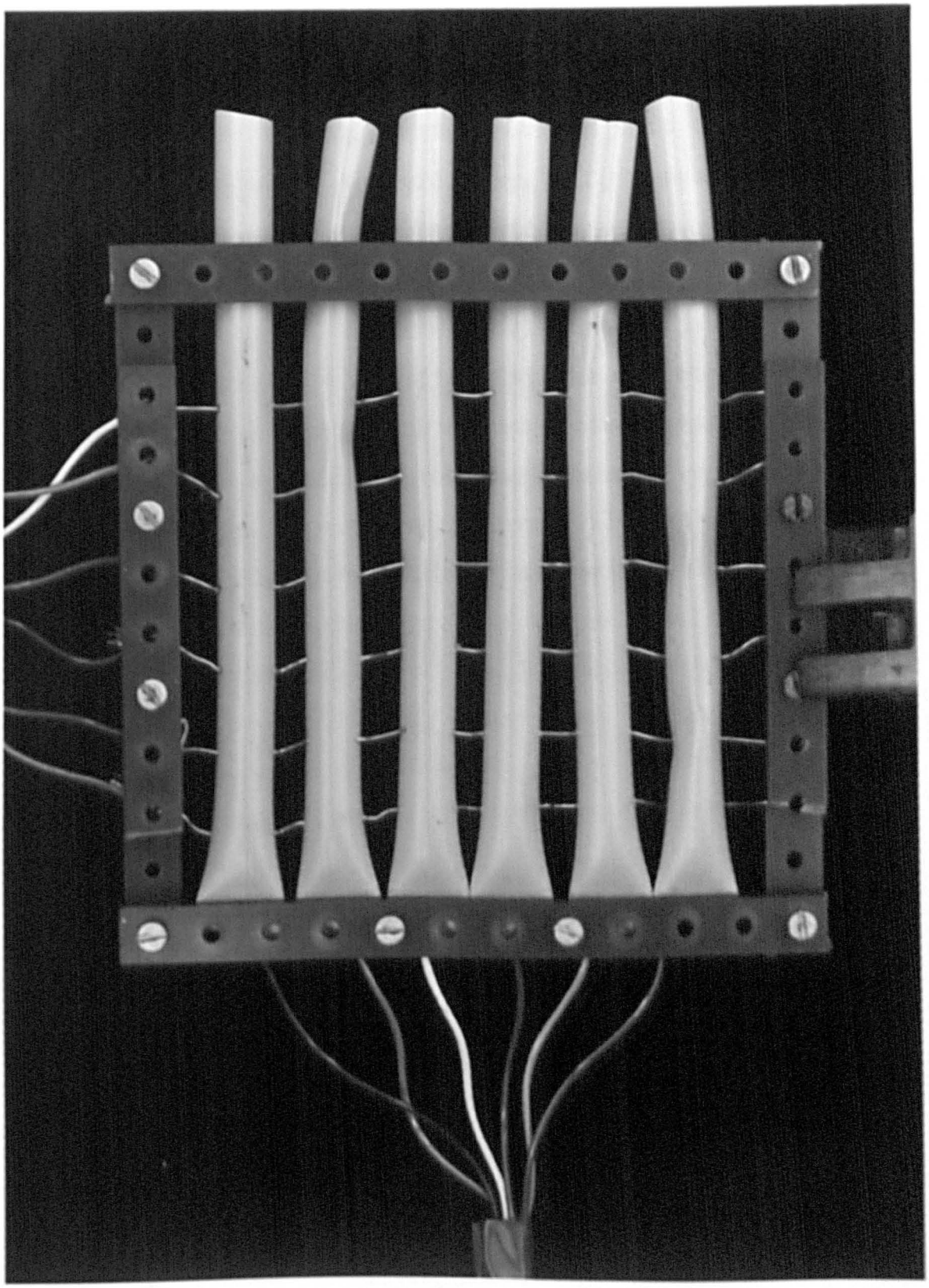


FIGURE 5.15.

variation at junctions situated far from the common electrode as at those near to it, provided that the impedances were adjusted individually. If, however, the impedances were changed by applying a D.C. voltage across all the junctions simultaneously, the impedances of the junctions near to the common electrode changed more than the others. This difficulty was overcome by employing small platinum contacts, connected externally, at both ends of the tube. A compound common electrode with inert contacts at regular intervals along the tube might be necessary for very long or narrow tubes.

Some experiments were performed to see how independently the impedances of the junctions could be adjusted. A serious difficulty arose when the impedance at one of the junctions was reduced. It will be appreciated that when currents are caused to flow through one junction other unwanted 'stray' currents flow through all the others. As the current required to break down the film is so much greater than that required to form one, these stray currents are sufficient to cause unwanted high impedance films to form at some of the junctions.

If the tubes are fitted with a means of emptying the electrolyte from them, the simplest solution to this 'stray current' problem is to remove the electrolyte from all tubes except the one which contains the impedance which is being changed. This eliminates stray currents. (MacKay and Ainsworth, 1963).

5 (7) Conclusions.

The variable impedance devices discussed in this chapter have properties which extend their potential application into areas different from those of the devices of Chapter III. The most important consequence of these properties is that they make possible the construction of large matrices of adjustable impedances in single units. They are also suitable as practical embodiments of the principle of effecting parallel switching by controlling the distribution of electrolyte.

They also possess a number of properties which will undoubtedly limit their usefulness. The unavoidable changes in the impedance of a cell for several hours after a D.C. voltage has been applied makes them unsuitable for most trial-and-error systems, and the large difference between forward and backward resistance makes circuits containing them much more difficult to design than if this were not the case.

These devices, however, are ideally suited to applications in which a uni-directional change in impedance is required.

REFERENCES, CHAPTER V.

Geel, W. Ch. van and Pistorius, C.A. (1956) "On the Residual Voltage with Electrolytic Capacitors."
Phillips Res. Report, vol. 11, pp 471-478.

MacKay, D.M. (1955) "The epistemological problem for automata",
Automata Studies, Princeton Univerty Press, pp. 235 - 251.

MacKay, D.M. (1956) "Towards an information flow model of human behaviour."
Brit. J. Psychol, vol. 47, pp 30- 43.

MacKay, D.M. (1961) "Information and Learning" in "Lernende Automata", Oldenburg, Munich, pp. 40- 49.

MacKay, D.M. and Ainsworth, W.A. (1961) "Improvements in or relating to automatic controlling devices and the like."
Brit. Prov. Pat. Spec. o12887/61.

MacKay, D.M. and Ainsworth, W.A. (1963) "Electrolytic Growth Processes with Applications to Self-Adjusting Automata" DAGK Symposium, Karlsruhe, in press.
(Appendix D. of this thesis.)

CHAPTER VI

CONDITIONAL PROBABILITY AND OTHER SYSTEMS

6 (1) Introduction.

The adaptive components described in the previous few chapters may be employed in any electrical circuit in which an integrator with a rather long time constant is required (Chapter VIII). A specific application of these devices is in a general type of machine which will be called an adaptive, conditional probability system.

Some years ago a conditional probability computer was constructed by Uttley (1955). This machine counted the coincidences of events in its environment, then computed the conditional probabilities of these events. If at any time this system were unable to receive signals from part of its environment it could, on the basis of these stored conditional probabilities and the signals it received from the rest of its environment, infer the events which were occurring in that part of the environment from which it could receive no signals.

A system of this kind may be used as a pattern classifier. Sets of signals, or patterns, together with other signals which represent the classes of these patterns are presented to the machine, and the machine calculates the conditional probabilities of these classes for each given input. At a later stage it uses these data to decide

which class a given pattern is most likely to belong to. A pattern recognising machine which operates in this manner has been built by Gamba et al (1961).

It will be shown in section (6.2) that many machines may be modelled in terms of a conditional probability system. The system described is different from that of Uttley in that it computes only the conditional probabilities of output signals for given input signals, not of the simultaneous occurrence of separate input signals. This, however, provides a convenient starting point for the design of an adaptive machine (Chapter VII). An adaptive machine A which is a realisation of this conditional probability system can obviously be trained to perform any task that a machine B (which can be modelled by the conditional probability system) can perform.

This system contains a matrix of conditional probabilities on which certain operations have to be performed in order to determine the behaviour of the machine. There are a number of other systems which perform the same operations on matrices of stored parameters, and so these may be regarded as conditional probability systems in a special sense. A number of these systems are described and their properties compared in the remainder of this chapter.

6 (2) Representation of a machine by conditional probabilities.

A machine is normally characterised by its transfer

function. This is a relation, or set of relations, between the signals which are applied to the machine, and the signals which are obtained from it. For a finite, discrete machine the transfer function may be represented by a set with a finite number of elements.

In a machine in which learning is taking place the transfer function will be continually changing. An observer, who had no information about the inside of a machine, might note that an input u_i produced an output v_j on n_1 occasions, and some other output on n_2 occasions. He might then say that the conditional probability of obtaining V_j given an input u_i was,

$$P(V_j / u_i) = \frac{n_1}{n_1 + n_2} \qquad (6.1)$$

In a similar manner, he could associate a conditional probability of this kind with each element of the transfer function of the machine.

Inputs to machines do not normally consist of one signal but of a combination or pattern of signals. A particular input signal may lead to one output when it occurs in one combination, and to another output when with another. An external observer would note the frequency of occurrence of each output with each input and draw up a matrix of conditional probabilities. He could do this for any machine with a pattern of input signals, even for

machines in which no learning was taking place.

Bayes' rule analysis.

The conditional probability of each output given an input pattern may be calculated from the conditional probabilities of each single element comprising the pattern if it is assumed that these elements are independent of each other. This has been shown by Minsky and Selfridge (1960) and by Maron (1962). The analysis in this section is based on the paper by Maron.

Consider a machine which has n input channels into each of which a binary signal is sent. The i^{th} input channel may contain a signal u_i or a signal $\bar{u}_i$, but not any other signal. The total input to the machine at any time is represented by the set (u_i).

Suppose the machine has a repertoire of m actions it can perform. Each of these is initiated by a signal in the output channel connected to a mechanism for producing this action. The j^{th} action is performed if the signal in the j^{th} output channel is v_j and not if the signal is $\bar{v}_j$. These are the only possibilities.

The conditional probability that the j^{th} action will be performed when the input is (u_i) is given by Bayes' rule as,

$$P(V_j / (u_i)) = \frac{P(V_j) \cdot P(u_i / V_j)}{P((u_i))} \quad (6.2)$$

It is shown in Appendix B that this may be written,

$$P(V_j / (u_i)) = \frac{1}{1 + \left[\frac{P(V_j)}{P(\bar{V}_j)}\right]^{n-1} \prod_{i=1}^{n} \frac{P(\bar{V}_j/u_i)}{P(V_j/u_i)}} \quad (6.3)$$

where $P(V_j / u_i)$ is the conditional probability that the output is v_j given an input signal u_i, and $P(\bar{V}_j / u_i)$ is the conditional probability that the output is not V_j given an input signal u_i. *

As this machine must either perform a given action or not perform it,

$$P(\bar{V}_j / u_i) = 1 - P(V_j / u_i) \quad (6.4)$$

If it may be assumed that the a priori probability of any given action $P(V_j)$ remains constant, it is possible to calculate the conditional probability of obtaining a signal in any output channel by means of equation (6.3), and (6.4) provided that the conditional probabilities of each output for the signal in each input channel are known.

6 (4) Interpretation as a machine.

The analysis of the previous section may be used as the basis for the design of a machine. Such a machine consists of n input channels and m output channels. Each connection between an input channel and an output channel has a 'weight' w_{ij} associated with it.

*If the values of $P(V_j / u_i)$ etc., remain constant, $P(V_j / (u_i))$ is really a conditional certainty.

As it is easier to build a machine which performs additions than multiplications, write

$$\log e \left\{ \frac{P(V_j / u_i)}{1 - P(V_j / u_i)} \right\} = L(V_j / u_i) \qquad (6.5)$$

Equation (6.3) then becomes,

$$P(V_j / (u_i)) = \frac{1}{1 + \left[\frac{P(V_j)}{P(\bar{V}_j)}\right]^{n-1} \exp - \sum_{i=1}^{n} L(V_j / u_i)} \qquad (6.6)$$

Figure (6.1) shows that $L(V_j / u_i)$ is a monotonically increasing function of $P(V_j / u_i)$ so that increases in $L(V_j / u_i)$ will increase $P(V_j / (u_i))$ (Equation 6.6) just as increases in $P(V_j / u_i)$ would have done (Equation 6.3).

If $L(V_j / u_i)$ is represented in the machine by $u_i \cdot w_{ij}$ (the product of a function representing the presence of a signal and a 'weight' representing a special kind of conditional probability) it is clear what operations a machine has to perform in order to calculate the conditional probability of a particular output given an input pattern. It needs to add together the weights of all the channels connecting active inputs to an output channel, and then to calculate the function given by equation (6.6).

For many applications it is not necessary for a machine to calculate explicitly the value of this function. It is only necessary for it to compute some measure of the

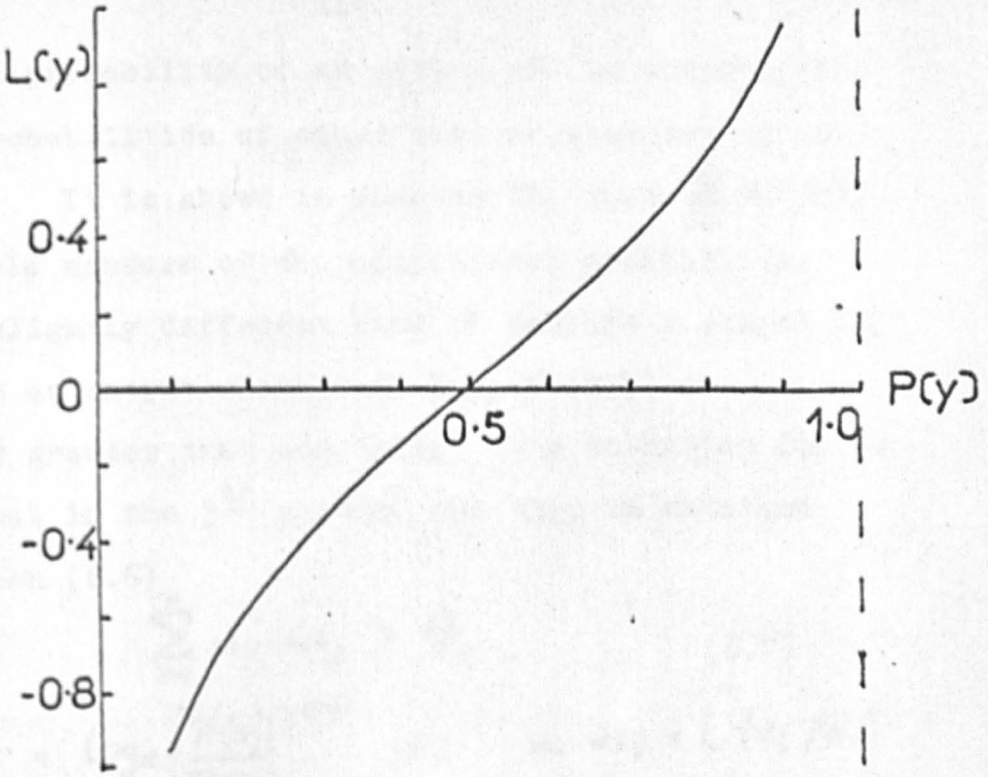


Figure 6.1.

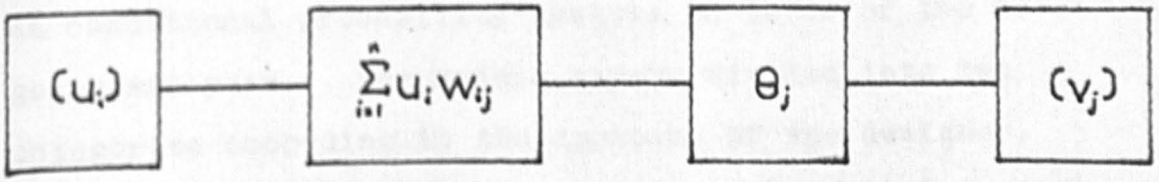


Figure 6.2.

conditional probability of an action and to compare this with the probabilities of other actions measured on the same scale. It is shown in chapter VII that $\sum_{i=1}^{n} u_i \cdot w_{ij}$ is a suitable measure of the conditional probability.

In a slightly different kind of machine a signal is produced in an output channel if $P(V_j / (u_i))$ is numerically greater than one half. The condition for an output signal in the $j^{\underline{th}}$ channel can then be obtained from equation (6.6)

$$\sum_{i=1}^{n} u_i \cdot w_{ij} > \theta_j \qquad (6.7)$$

where $\theta_j = \log_e \left[\frac{P(V_j)}{P(\bar{V}_j)}\right]^{n-1}$ and $u_i \cdot w_{ij} = L(V_j / u_i)$.

The operations which such a machine needs to perform are shown diagrammatically in Figure (6.2).

6 (5) Some models of conditional probability systems.

A number of mathematical models and physical machines have been developed which may be interpreted as conditional probability systems in terms of the foregoing analysis. The models may be divided into two categories according to the approach of the designer.

Firstly there are the nets of artificial neurons. Neurons, or nerve cells, are generally supposed to be the main information processing units in nervous systems. An artificial neuron is a mathematical or electronic device which purports to have the same gross functional properties as biological neurons. Electronic models have

been built by Harmon (1959) and others. These neurons have a number of input channels which receive signals. These signals are added together, and if this sum exceeds a certain threshold value a signal appears in the output channel. In some of these models a delay is introduced between the reception of the input signals and the appearance of an output. Other features, such as a refractory period during which there can be no output no matter what input signals are applied, are also incorporated into some of these models.

One of the first attempts to describe nets of neurons by means of a mathematical model was by McCulloch and Pitts (1943). Numbers of artificial neurons were considered to be connected together by "synapses" to form nets. The 'synapses' were given weights which modified the strengths of the signals passing through them. In the subsequent analysis of nets of this type McCulloch and his collaborators have shown that the logical properties of these nets need not be affected by certain changes in thresholds and synaptic weights (McCulloch et al, 1962). On the other hand the threshold values and the synaptic weights may be adjusted to bring about definite changes in the logical functions computed by the net.

The properties of a net of artificial neurons connected together randomly have been investigated by Rochester et al

(1956) by simulation on a digital computer. They found that unless inhibitory connections (synapses with negative weights) were introduced the activity in their net increased continuously. Beurle (1956), studying the properties of randomly connected nets analytically, found that they were capable of supporting various activities such as plane and spherical waves. Allanson (1956) has shown how the values of parameters of randomly connected nets affect the activity in the nets, and how some of these activities might be linked to the behaviour of nervous systems.

Willis (1959) and Caianiello (1960) have considered nets of artificial neurons with adaptive connections. An additional problem arises here as two distinct processes are taking place; the primary computing activity of the neurons and the changes in this activity brought about by changes in synaptic weights. It is considered by Caianiello to be possible to treat these processes separately as they are supposed to take place with very different time constants.

The second approach which may involve designing a conditional probability system is to build a pattern recognising machine. One possible mode of operation of such a machine was explained in the introduction. Taylor (1959), for example, has constructed a machine which performs a similar operation, a particular feature of which

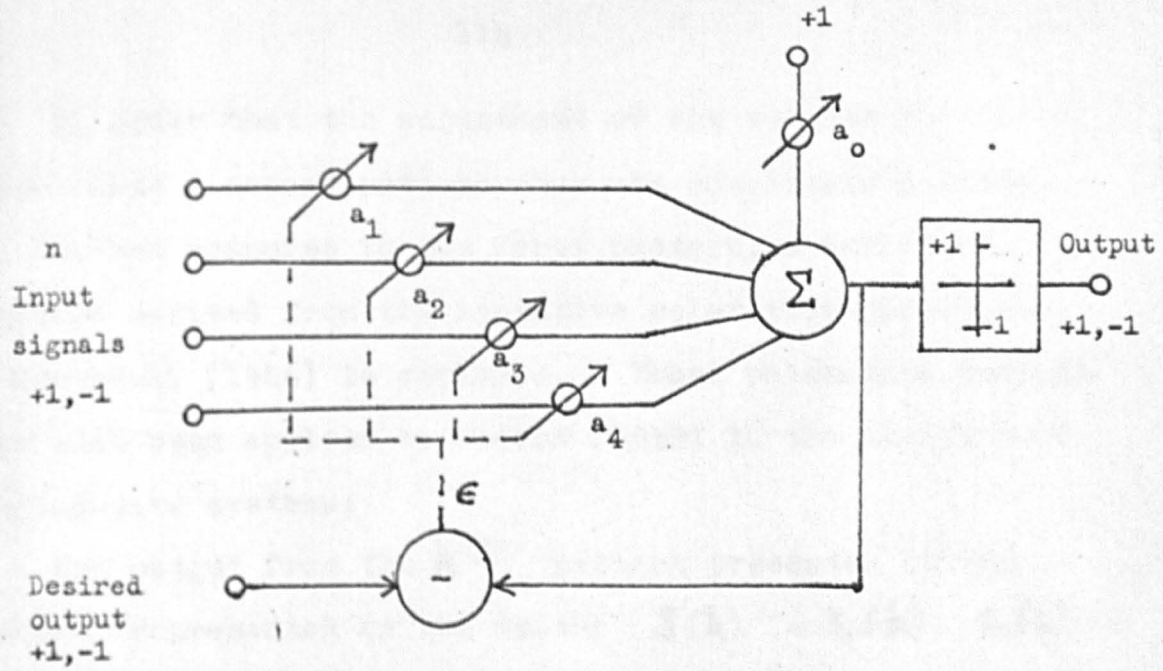
is that it can be trained to discriminate between patterns, and so may be considered to be an adaptive probability computer.

Uttley (1954) has synthesised these various approaches as he postulated a theory of pattern recognition by the nervous system in terms of the design and construction of a conditional probability computer. (Uttley, 1955). More recently Gamba (1961) has built a machine called P.A.P.A. which can be trained to discriminate between patterns by the calculation of conditional probabilities. This machine appears to be capable of forming certain generalisations.

Recent years have seen a number of additions to the field of adaptive automata and pattern recognising systems. Some of these will be described in greater detail in the next few sections.

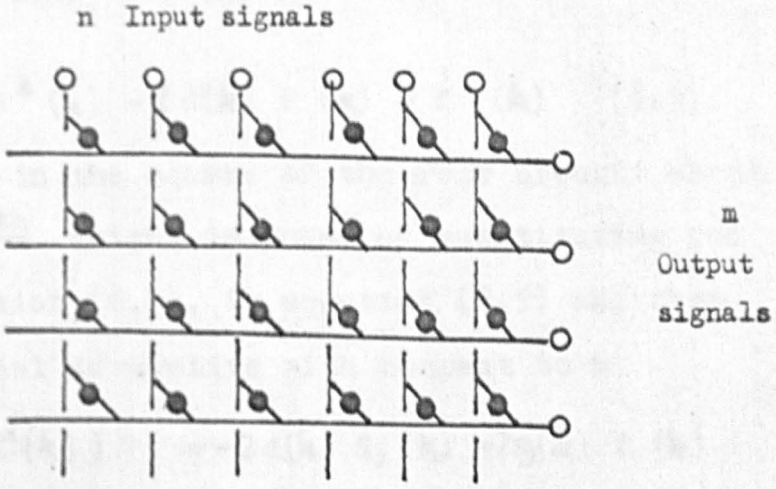
6 (6) "Adaline neuron"

The "adaline neuron" designed by Widrow (1960, 1962) is a machine which performs the operations shown in Figure (6.2). It consists of n input channels, which may carry a signal + 1 or - 1, connected to a quantizer via variable weights a_1 to a_n . (Figure 6.3). The threshold is adjusted by means of the weight a_0 which is always connected to a + 1 input. Patterns of signals (+1, -1) are presented to the input channels and the weights are adjusted so that the desired output from the quantizer is obtained.



"Adaline Neuron"

Figure 6.3.



"Learning Matrix"

Figure 6.4.

In order that the adjustment of the weights to accommodate a second pattern does not completely destroy the desired response to the first pattern, a method of adaption derived from the iterative relaxation procedures of Southwell (1946) is employed. These relaxation methods have also been applied by Widrow (1959) to the analysis of sampled-data systems.

The output from the k^{th} pattern presented to the machine, represented by the vector $S(k) = s_1(k) \ldots s_n(k)$ may be written, before quantization, as

$$f(k) = a_0 + a_1 s_1(k) + a_2 s_2(k) + \ldots + a_n s_n(k) \quad (6.8)$$

If the desired output is $d(k)$ then the error $\varepsilon(k)$ will be,

$$\varepsilon(k) = d(k) - f(k)$$

or $$\varepsilon^2(k) = d^2(k) - 2\,d(k)\,f(k) - f^2(k) \quad (6.9).$$

The change in the square of the error brought about by altering the j^{th} weight is found by substituting for $f(k)$ from equation (6.8), in equation (6.9) and then taking the partial derivative with respect to a

$$\frac{\partial}{\partial a_j}\left(\varepsilon^2(k)\right) = -2\,d(k)\,s_j(k) + 2 s_j(k)\,f(k)$$
$$= -2\,s_j(k)\,\varepsilon(k) \quad (6.10).$$

Each element of the vector $S(k)$ has the same absolute magnitude so adjustment of any weight by a given

amount will affect the total error by the same amount. The method used by Widrow for an n-bit input pattern is to change each weight in the direction which reduces the error. If all weights are changed by the same quantity each individual change reduces the total error by $1 / (n+1)$, and the final error after changing n weights plus the threshold a_o is zero.

A purely automatic 'adaline neuron' has been built by Widrow and Hoff (1962). The negative input signals (-1) are A.C. signals 180° out of phase with the positive signals (+1). 'Memistors' (v. Chapter III) are employed to make the system adaptable, and negative weights are produced by connecting these devices in bridge-type circuits.

Hoff (1962) has developed methods of analysis of these circuits. These adaptive switching circuits can only be used to classify a sub-set of the 2^n possible input patterns into desired categories.

For example, it is shown by Hoff to be impossible for a single adaline neuron to map a pattern and its complement into the same category. It is possible, however, for networks of adaline neurons to be trained to perform such operations. Hoff has also shown that adaline neurons may be trained to perform certain generalisations such as reacting similarly to patterns which are rotated through 90° with respect to each other.

6 (7) "Learning Matrix."

A proposal for a system which may be interpreted as a physical realisation of a conditional probability computer was given by Steinbuch (1960). This was called a "learning matrix". It consists of a matrix of horizontal and vertical wires as illustrated in Figure 6.4. A pattern of signals is presented to the column wires and a new pattern, which depends on the transmission coefficients of the junctions, is obtained from the row wires. The learning takes place by the adjustment of the transmission coefficients of the junctions. This system is considered by Steinbuch to be analogous to the "conditioned reflexes" of animals first investigated in the classical researches of Pavlov (1927).

The method of adjustment given by Steinbuch is, in principle, very simple. The pattern of input signals determines a number of columns and the desired output determines a certain row. The transmission coefficients of the junctions between these columns and this row are then increased. Whenever this pattern of signals is presented in the future the signals on this particular output line should then be larger than on any of the others, and this can be detected. This procedure can be repeated for a number of patterns of signals so that each can be made to switch to a particular output line.

Patterns of signals which were not present in the

set presented during the 'learning' phase will also select output lines. The output selected by one of these is the row which the pattern in the training set which differed from this pattern by the smallest Hamming distance (Hamming, 1950) would have selected. This enables the system to be used as an error-corrector.

"Learning matrices" have been constructed and described by Steinbuch et al (1961). A number of processes have been investigated as possibilities for realising the variable transmission coefficients of the junctions. These may be classified as magnetic, breakdown of condensers, and electrochemical. The magnetic devices, in which the direction of magnetisation of a core may be switched from one state to another, have proved the most successful, and a number of matrices have been built with these devices. When electrochemical variable-impedance devices were employed difficulties caused by the 'stray paths' were encountered (v. Chapters V and VII).

6 (8) "Perceptron."

A number of systems which have become known as "perceptrons" have been built by Rosenblatt and his associates (Rosenblatt, 1960; Block et al, 1962). These machines do not have the regular structure of the "adaline neurons" and "learning matrices" as they are intended primarily as possible models of nervous systems. They do, however, perform the same logical operations as these

machines. Moreover, they could be manufactured from the same adaptive components.

A simple perceptron consists of three layers (Figure 6.5). The sensory units S are transducers which convert the input patterns of light, or sound, into electrical signals. These electrical signals are passed through a network of random connections to the association layer. The signals from these A-units then pass through a network of variable weights to the response units, R. The part of the machine to the right of the A-units (Figure 6.5) is very similar to an adaline neuron.

The operation of the machine is as follows. A stimulus is presented which turns on some of the S-units. These subsequently cause some of the A-units to become activated, and a response is obtained from those R-units whose thresholds are exceeded. The desired response is obtained for a given stimulus pattern by varying the weights of the connections between the A- and R- units by means of an "error correction" procedure. A number of these training routines have been studied by Rosenblatt (1960a). In the simplest of these, if the response is correct no changes are made but if the response is erroneous all weights associated with the active A-units are decremented.

The random net between the S-units and the A-units is a useful device as it makes possible the classification

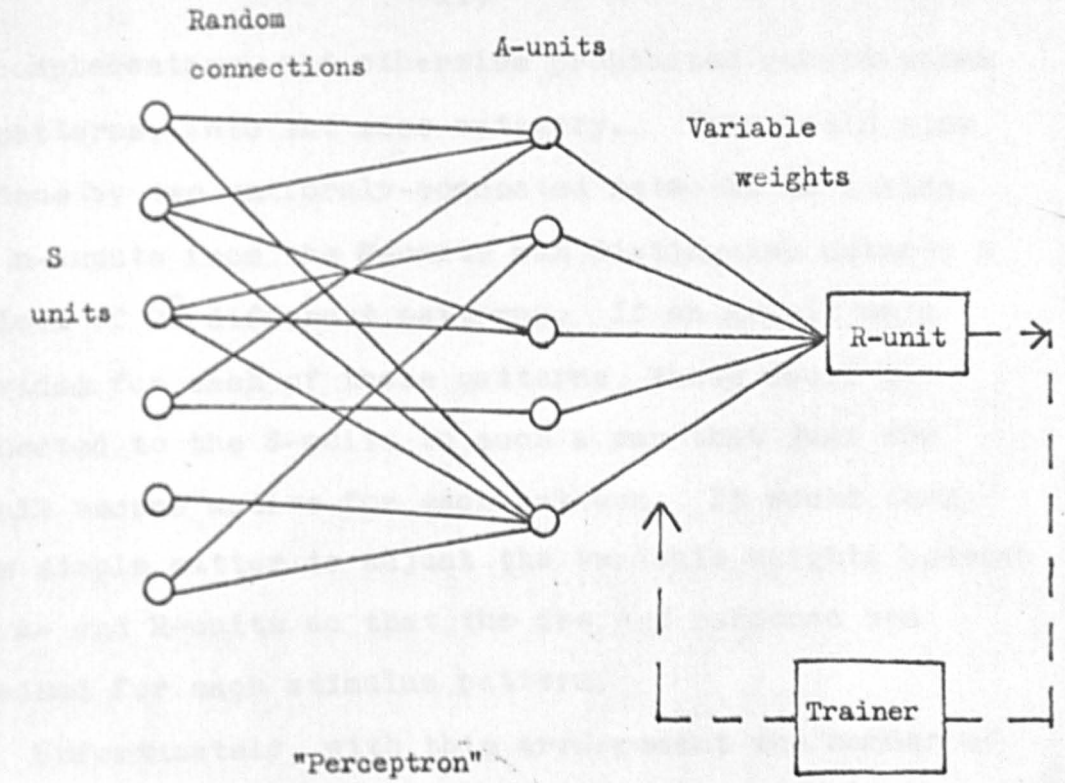


Figure 6.5.

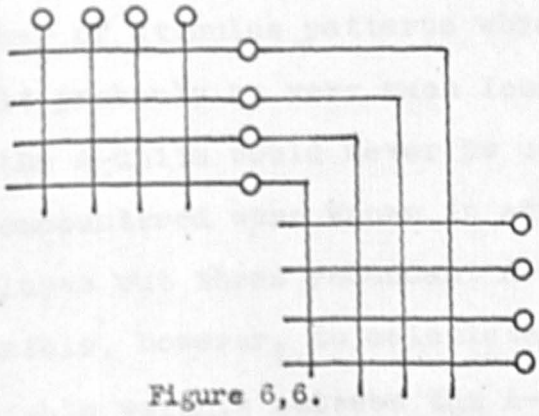


Figure 6,6.

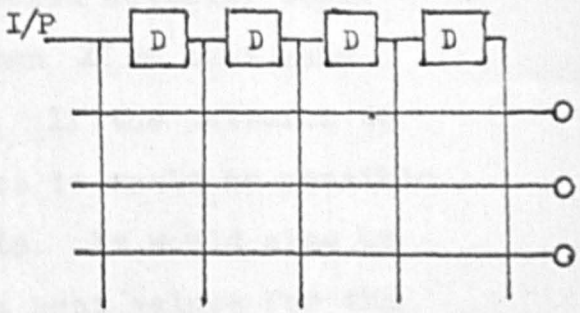


Figure 6.7.

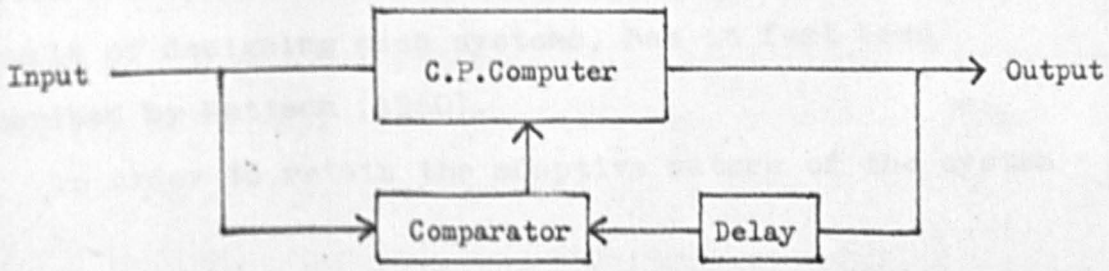


Figure 6.8.

of complementary, and otherwise prohibited combinations of patterns, into the same category. This could also be done by two uniformly-connected networks in series. The n-inputs from the S-units can distinguish between a maximum of 2^n different patterns. If an A-unit were provided for each of these patterns, these could be connected to the S-units in such a way that just one A-unit became active for each pattern. It would then be a simple matter to adjust the variable weights between the A- and R-units so that the desired response was obtained for each stimulus pattern.

Unfortunately, with this arrangement the number of A-units would increase astronomically with the number of S-units and such machines would be uneconomical for practical applications. For a given application the number of stimulus patterns which would actually occur would probably be very much less than 2^n so that many of the A-units would never be used. If the patterns to be encountered were known in advance it would be possible to leave out these redundant A-units. It would also be possible, however, to calculate the best values for the variable weights between the A- and R-units, so a fixed system could be designed and built. A computer programme capable of designing such systems, has in fact been described by Mattson (1960).

In order to retain the adaptive nature of the system

and still keep the number of A-units within practical bounds, the connections between the S- and A- units could be made variable. This would, however, introduce further complications into the training procedure. The employment of a random net seems to be a useful compromise between these difficulties.

The original perceptrons (Rosenblatt, 1960b) contained motor-driven potentiometers as their adaptive components. Later versions, such as Tobermory, are being built which employ more economical and small-scale integrators, such as magnetostrictive devices, as their variable weights (Nagy, 1962).

6 (9) Problems with adaptive systems.

There remains a number of problems which, although they have been partially solved by the designers of the adaline neurons and perceptrons, still require more analysis and experiment. It is thought that more work should be done on the development of methods which will enable a machine designer to guarantee that a system will converge to some desired behaviour within a finite time. Although adjustment of synaptic weights in the direction which will increase the conditional probability (this direction may be determined from equation (6.7)) will enable the system to 'learn' one input / output relationship it is by no means guaranteed that learning another will not upset the first relationship,. Interference of this type between successive adjustments

of weights may make it impossible for a particular set of relationships to ever be learned. Both Widrow (1962) and Rosenblatt (1960a) have shown convergence is possible for some adaptive systems and training procedures.

The next problem is to develop methods of choosing between various training procedures. It is normally desired that the set of patterns should be learned as quickly as possible. Some training procedures might be rapid, while others might be better able to cope with "noisy" patterns, or with a trainer which sometimes made mistakes.

Another problem arises when an extra layer of adaptive connections and threshold elements is included in the system. As an example consider two learning matrices connected in series so that the output signals from the rows of the first matrix, after passing through threshold units, become the input signals to the columns of the second matrix (Figure 6.6). Suppose this system gives an undesired output for a particular pattern. Should the transmission coefficients of the junctions in the first matrix, or the second, or both be adjusted? Probably all should be altered but this might mean spoiling a perfectly good set of weights in one of the matrices. In such systems the interference will be much greater and the training period longer.

6 (10) Time dependent systems

In addition to the automata so far discussed a number

of systems have been proposed or constructed which fall into the same general category but in which time delays play an important role. These have purposely not been considered until now.

There is a simple device for transforming a pattern in time, such as a sequence of pulses, into a pattern in space. This consists of a series of delays as shown in Figure (6.7). This device has been used by Uttley (1955). The points between successive delays may be connected to the inputs of a learning matrix or adaline neuron and then, after a few appropriate modifications, the system can be trained to discriminate between patterns in time.

The addition of a number of delay lines to a conditional probability computer may be made to produce a self-organising system capable of prediction (Figure 6.8). The machine is required to predict what the input will be at time (t+1) given the input at time t. The output at time t is fed through a bank of parallel delay lines, each having unit time delay, to a comparator.

At (t+1) the output of these delay lines is compared with the input to the system. As a result of this evaluation the 'weights' in the computer are adjusted automatically so as to increase the probability that the 'error' signal from the comparator should be zero in future. Provided that there is some order in the presentation of

the patterns the system should eventually learn to predict the next pattern. A system of this kind has been proposed by Maron (1962). With more delay lines and more adaptive components it is possible to extend the system so that the prediction of the pattern at (t+1) depends on the patterns at (t-1), (t-2), etc. as well as on that at time t. Another predictor which works by extrapolation has been reported by Gabor et al (1960).

Time delays have been used by Block et al (1962) in a four layer perceptron which, it is claimed, can learn to distinguish between patterns automatically. The perceptron has two layers of A-units connected by variable resistance pathways and delay lines. Patterns are presented at regular intervals equal to the delay in these lines. If a unit in the first association layer is active at the same time as one in the second the resistance of the path between them is decreased. The sequence of the patterns presented is such that each pattern is more likely to be the same as the previous one than not. This perceptron takes advantage of this redundancy in the environment and so learns to discriminate between different patterns without the intervention of a human trainer. In order to get a desired response from a particular pattern it is still, of course, necessary to train the connections between the second layer of A-units and the R-units with a human operator.

A class of active delay lines, called neuristors, has been described by Crane (1960). These devices transmit signals without any decay in signal strength in a similar way to the axon of a nerve cell. Crane has shown that networks of these devices can perform all the logical computing functions. If these units were connected together with variable impedance devices adaptive systems would result which could be trained in much the same way as perceptrons and adaline neurons.

Another method of producing adaptive behaviour is possible with neuristors. MacKay has drawn attention to the possibility of storing information in the nervous system (MacKay and McCulloch, 1952; MacKay, 1954) and in self-organising systems (MacKay, 1962) by the adjustment of the delay in transmission between units. He has suggested that if neuristors could be produced in which the delay time per unit length was adjustable, these might provide an attractive alternative to variable impedance devices in systems sensitive to the coincidences of impulses.

REFERENCES, CHAPTER VI

Allanson, J.T. (1956) "Some Properties of a Randomly Connected Network," Third London Symposium on Information Theory, Butterworth, London, pp 303 - 313.

Beurle, R.L. (1956) "Activity in a Block of Cells Capable of Regenerating Pulses." Phil. Trans. Roy. Soc., B 204, p. 55.

Block, H.D. (1962) "The Perceptron : A Model for Brain Function." Rev. Modern Physics, vol. 34, pp. 123-135.

Block, H.D., Knight, B.W. and Rosenblatt, F, (1962) Rev. Modern Physics, vol. 34, pp 135-142.

Caianiello, E.R. (1961) "Outline of a theory of Thought Processes and Thinking Machines." J. Theoretical Biology, vol. 2, pp 49-52.

Crane, H.D. (1960) "Neuristor Studies." T.R. No. 1506-2, Stanford Electronics Lab., Stanford, California.

Gabor, D., Wilby, W.P.L., and Woodcock, R (1960) "A Self-Optimizing Non-Linear Filter, Predictor and Simulator". Fourth London Symposium on Information Theory,. Butterworths, pp. 348 -352.

Gamba, A. et al (1961) "Further Experiments with P.A.P.A." Supplement to Nuovo Cimento, vol. 20, no. 2, pp. 112-221.

Hamming, R.W. (1950) "Error Detecting and Error Correcting Codes." Bell Systems Tech. Journal, vol. 26, no. 2, pp. 147-160.

REFERENCES, CHAPTER VI (Continued)

Harmon, L.D. (1959) "Artificial Neuron."
Science, vol. 129, no. 3354, pp 962-3.

Hoff, M.E. (1962) "Learning Phenomena in Networks of Adaptive Switching Circuits". T.R. No. 1554-1, Stanford Electronics Laboratory.

MacKay, D.M. and McCulloch, W.S. (1952) "The Limiting Information Capacity of a Neuronal Link."
Bull. Math. Biophysics, vol. 14, pp 127-135.

MacKay, D.M. (1954) "Operational Aspects of Some Fundamental Concepts of Human Communication."
Synthese, vol. 9, pp. 182-198.

MacKay, D.M. (1961) "Information and Learning" in "Lernende Automaten", Oldenburg, Munich, pp. 40-49.

MacKay, D.M. (1962) "Self-Organization in the Time Domain." in "Self-Organizing Systems, 1962". Spartan Books, Washington, pp. 37-48.

Maron, M.E. (1962) "Design Principles for an Intelligent Machine."
I.R.E. Trans. on Information Theory, vol IT-8, no. 5, pp. 179-185.

Mattson, R.L. (1960) "An Approach to Pattern Recognition using Linear Threshold Devices".
T.R. LMSD -702680.

Minsky, M. and Selfridge, O.G. (1960) "Learning in Random Nets", Fourth London Symposium on Information Theory, Butterworths, London, p. 338.

Nagy, G. (1962) "Analogue Memory Mechanisms for Neural Nets." Cognitive Systems Research Program. Report No. 3, Cornell University, Ithaca, New York.

REFERENCES, CHAPTER VI (Continued)

Pavlov, I.P. (1927) "Conditioned Reflexes."
Oxford University Press, London.

Rochester, N., Holland, J.H., Haibt, L.H., and Duda, W.L. (1956).
"Tests on a Cell Assembly Theory of the Action of the Brain,"
I.R.E. Trans., IT - 2, pp. 80-93.

Rosenblatt, F. (1960a) "On the convergence of reinforcement procedures in simple perceptrons."
Report No. VG -1196-G-4, Cornell Aeronautical Laboratory, Buffalo, New York.

Rosenblatt, F. (1960b) "Mark I Perceptron Operators' Manual."
Report No. VG-1196-G-5, Cornell Aeronautical Laboratory, Buffalo, New York.

Southwell, R.V. (1946) "Relaxation Methods."
Oxford University Press.

Steinbuch, K. (1961) "Die Lernmatrix".
Kybernetick, vol. 1, no. 1, pp. 36-45.

Steinbuch, K. et al (1961) "Schaltungen mit der Lernmatrix."
in "Lernende Automaten" Oldenburg, Munich, pp. 63-100.

Taylor, W.K. (1959) "Pattern Recognition by Means of Automatic Analogue Apparatus" Proc. I.E.E. pp. 198-209.

Uttley, A.M. (1954) "The Classification of Signals in the Nervous System."
E.E.G. Clin. Neurophysical, vol.6, pp.479-494.

Uttley, A.M. (1955) "The Conditional Probability of Signals in the Nervous System."
R.R.E. Memo. No. 1109, Ministry of Supply.

REFERENCES, CHAPTER VI (Continued)

Uttley, A.M. (1956) "Conditional Probability Machines and Conditioned Reflexes".
Automata Studies, Princeton University Press.
pp. 253-285.

Widrow, B. (1960a) "Adaptive Sampled-Data Systems".
T.R. No. 2104-1, Stanford Electronics Laboratory.

Widrow, B. (1960b) "An Adaptive 'Adaline' Neuron using Chemical 'Memistors'".
T.R. No. 1553-2, Stanford Electronics Laboratory.

Widrow, B. (1962) "Generalization and Information Storage in Networks of Adaline Neurons" in "Self-Organizing Systems 1962".
Spartan Books, Washington, pp. 435-461.

Willis, D.G. (1959) "Plastic Neurons as Memory Elements" in "Proc. Internat. Conf. on Information Processing."
J.6.7., Unesco House, Paris.

CHAPTER VII

A PHYSICAL REALISATION OF AN ADAPTIVE SYSTEM

7 (1) Introduction

In order to fulfil the original purpose of this investigation an adaptive system has been constructed which employs the electrolytic variable impedance devices of Chapter III. This system is an adaptive conditional probability computer in the sense of the previous chapter. It is similar to the "learning matrix" in its electrical circuit and the training procedure used is similar to one of those employed with "perceptrons". There are, however, a number of novel features incorporated into the design which have been necessitated by the use of electrolytic variable resistors as the adaptive compnnents.

7 (2) Direction of weight changes.

The design of the system is based on equation (6.6). The machine is a system of n input channels, each of which may carry a signal u or $\bar{u}$, connected by adjustable pathways to m output channels. For a given set of signals in the input channels (u_i), the machine is required, when the training procedure is completed, to produce a signal V in the output channel desired by the trainer and a signal $\bar{V}$ in all the other output channels. It does this by computing a measure of the conditional probability $P(V_j / (u_i))$ for each output channel, and then choosing

the channel which has the greatest probability of being correct.

This may be expressed as choosing j for which

$$P(V_j / u_i) > P(V_k / u_i) \quad \text{for all } k \neq j. \tag{7.1}$$

In terms of equation (6.6) this becomes,

$$\frac{1}{1 + \left[\frac{P(V_j)}{P(\bar{V}_j)}\right]^{n-1} e^{-\sum_{i=1}^{n} L(V_j/u_i)}} > \frac{1}{1 + \left[\frac{P(V_k)}{P(\bar{V}_k)}\right]^{n-1} e^{-\sum_{i=1}^{n} L(V_k/u_i)}}. \tag{7.2}$$

Suppose that each output is a priori equally likely, so that,

$$P(V_j) = P(V_k) = 1/m$$

and

$$P(\bar{V}_j) = P(\bar{V}_k) = (m-1)/m.$$

Equation (7.2) then becomes,

$$1 + (m-1)^{n-1} \exp\{-\Sigma' L(V_k/u_i)\} > 1 + (m-1)^{n-1} \exp\{-\Sigma' L(V_j/u_i)\}$$

or

$$\sum_{i=1}^{n} L(V_j/u_i) > \sum_{i=1}^{n} L(V_k/u_i) \quad \text{for all } k \neq j. \tag{7.3}$$

$L(v_j / u_i)$ is a measure of the probability that a signal in channel i should lead to a signal in the output channel j. This may be computed by the product of the input signal u and the 'weight', or conductance, of the path which connects these channels, W_{ij}.

$$L(V_j / u) = u_i \cdot w_{ij} \tag{7.4}$$

It is now clear how the weights should be adjusted in order to change the conditional probabilities. If the machine makes a correct response all the weights w_{ij} of pathways which connect input channels containing signals to the desired output should be increased relative to all

the other w_{ij}.

If the signals in the input channels are allowed to be positive or negative the situation is a little more complicated. The weights must then be altered by an amount proportional to,

$$Z \cdot \frac{w_{ij}}{|w_{ij}|} \cdot u_i \cdot v_j \qquad (7.5)$$

where Z, the training operator, is + 1 if the desired response is obtained and - 1 otherwise, u_i is the input to channel i (= 1 or - 1), and V_j is +1 if the j^{th} output channel has the maximum probability and 0 otherwise.

In order to avoid complications caused by negative input channels and negative weights each input channel was duplicated so that a signal u_i produced a 1 in one channel and a 0 in the other and $\bar{u}_i$ produced 0 in the first channel and 1 in the second. Weights which were always positive could then be employed. These are represented by w_{ij} if a signal u_i is received and $\bar{w}_{ij}$ if a signal $\bar{u}_i$. In order that the training procedure should have the desired result the weights which must be adjusted are the ones for which $u_i \cdot v_j$ is not zero, and the direction of change is given directly by the training operator, Z.

It is not sufficient to use signals 1 and 0 and a single channel between i and j. In this case if the input signal is 0, the product $u_i \cdot w_{ij}$ on which the decisions are based is 0 whatever the value of w_{ij}. The machine is thus not making full use of its past experience.

7 (3) Magnitude of the weight changes.

The training procedure employed with the system may be summarised as follows,

Desired output obtained,

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w \qquad (7.6a)$$

Undesired output obtained,

$$w_{ij}(t+1) = w_{ij}(t) - c\Delta w \qquad (7.6b)$$

where Δw and c are constants.

The values of Δw and c will affect the speed with which the system can be trained to give a correct set of responses. If Δw is large w_{ij} may be changed quickly. This may mean that the optimum value for w_{ij} is rapidly reached or it may be that this will increase the number of incorrect responses.

The mean change in w_{ij} for a system in which this reinforcement procedure is employed may be obtained as follows. Suppose that after r sets of inputs (patterns) have been presented the probability of obtaining a correct response is P (r), the mean change in any weight will be,

$$w(t+r) = w(t) + \frac{r\Delta w}{m} \sum_{r}' P(r) - \frac{rc\Delta w}{m} \sum_{r}' (1-P(r)). \qquad (7.7)$$

assuming that there are m outputs and each is chosen equally often.

In an ideal machine in which each output is a priori

as equally likely as any other it would be expected that one correct response would be obtained while the first m patterns were presented. The probability of a correct response is,

$$P(m) = {}^{1}/_{m}$$

and in an ideal machine the correct response would be learned and remembered.

For each of the next m trials (presentations of a pattern) there would be a ${}^{1}/_{m}$ chance of obtaining this learned response and an (m-1)/m chance of obtaining one of the others. For each of these others there would be a 1/(m-1) chance of obtaining the desired response. Hence,

$$P(r) = \frac{r}{m^2} \qquad (7.8)$$

The ideal machine should learn the correct set of responses in about m trials.

Substituting equation (7.8) in equation (7.7),

$$w(t+r) - w(t) = \frac{r^2 \Delta w}{m} \left\{ \frac{(1+c)(r+1)}{2m^2} - c \right\} \qquad (7.9)$$

as $\sum_{r} P(r) = \frac{r(r+1)}{2m^2}$

If the patterns are learned in m^2 trials the mean change in the weights will be,

$$w(t+m^2) - w(t) = \frac{m \Delta w}{2} \left\{ (1+c)(m^2+1) - 2cm^2 \right\}$$

$$\text{or } w(t+m^2) - w(t) = \frac{m^3 \Delta w \ (1-c)}{2} \qquad (7.10)$$

as $m^2 \gg 1$.

In the absence of a general theory of adaptive systems the values of Δw and c were chosen according to the following considerations.

Suppose that the highest practical value attainable by a weight is w_1 , the lowest w_2 , and the mean value w_3 . The sum of the signals to the j^{th} output channel at a time t may then be expressed as,

$$y_j (t) = n \ w_3 + \epsilon , \qquad (7.11)$$

where n is the number of input channels, and ϵ is an arbitary quantity.

If this output is chosen, and it is the desired one, the reinforcement procedure will increase this to,

$$y_j (t+1) = n(w_3 + \Delta w) + \epsilon . \qquad (7.12).$$

Suppose now a pattern of signals differing by one Hamming distance (v. Chapter VI) is presented, the output at the j^{th} channel will now be at least,

$$y_j' (t+1) \geq (n-1)(w_3 + \Delta w) + w_2 + \epsilon , \qquad (7.13)$$

where w_2 is the minimum value of the weight.

As this is a new pattern it should be capable of bringing forth a new response. It may only do this if,

$$y_j' (t+1) < y_j (t) , \qquad (7.14)$$

provided that no other weights have been altered.

From equations (7.10), (7.12) and (7.13),

$$(n-1)\, w_3 + W_2 + (n-1)\Delta W < n\, w_3 ,$$

$$\text{or } \Delta w < \frac{w_3 - w_2}{n-1}$$

Writing $w_3 = \frac{w_1 + w_2}{2}$ this becomes,

$$\Delta w < \frac{w_1 - w_2}{2\,(n-1)} \qquad (7.15)$$

Suppose again that the output channel j is chosen but that this is not now the required response. The reinforcement procedure will now change $y_j\,(t)$ to,

$$y_j\,(t+1) = n\,(w_3 - c\,\Delta w) + \epsilon , \qquad (7.16).$$

If this response was not the desired one a second presentation of the same pattern should produce a different response. The output channel k with the second largest signal at the first presentation would have carried a signal less than $(n\, w_3 + \epsilon)$, approximately,

$$y_k(t) = (n-1)\, w_3 + w_2 + \epsilon . \qquad (7.17)$$

If this is the greatest signal at t + 1,

$$y_k\,(t+1) > y_j\,(t+1)$$

$$\text{or } w_2 > w_3 - n\, c\, \Delta w$$

$$c.\Delta w > \frac{w_1 - w_2}{2\,n} \qquad (7.18)$$

Another relationship may be obtained as follows. A pattern of signals is presented and this obtains the

desired response, so after reinforcement the signal strength is given by equation (7.12). Now suppose a new pattern, differing by one Hamming distance, is presented and this also gives the same response, but this is not the desired one for this new pattern. The output at (t+1) with this new pattern will be approximately,

$$y'_j\ (t+1) = (n-1)(w_3 + \Delta w) + W_2 + \epsilon \qquad (7.19.$$

and after reinforcement at (t+2) this becomes,

$$y'_j\ (t+2) = (n-1)(W_3 + \Delta w - c\Delta w) + w_2 + \epsilon \qquad (7.20)$$

(w_2 is the minimum conductance).

Now if the original pattern is reapplied,

$$y_j\ (t+2) = (n-1)(w_3 + \Delta w - c\Delta w) + w_3 + \Delta w + \epsilon \qquad (7.21)$$

This should give the same result as before, so

$$y_j\ (t+2) \geqslant y_j\ (t)$$

From equations (7.12) and (7.21),

$$n\, w_3 + n\Delta w - (n-1)\, c\Delta w + \epsilon \geqslant n\, w_3 + \epsilon,$$

$$\text{or} \quad c \leqslant \frac{n}{n-1} \qquad (7.22)$$

These inequalities (7.15) (7.18) (7.22) are shown plotted as Δw versus c in Figure (7.1). This shows the region for which values of Δw and c are permissible.

Alternatively as $n / (n-1) \geqslant 1$ for $n \geqslant 1$ it is possible to set $c = 1$ (Equation (7.22) and choose Δw such that,

$$\frac{w_1 - w_2}{2n} \leqslant \Delta w \leqslant \frac{w_1 - w_2}{2(n-1)} \qquad (7.23)$$

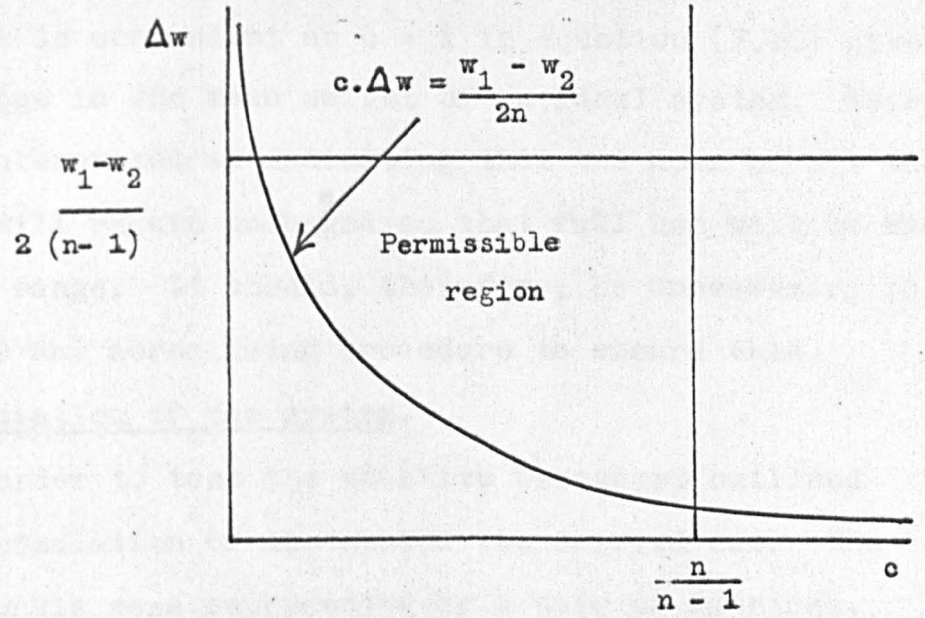


Figure 7.1.

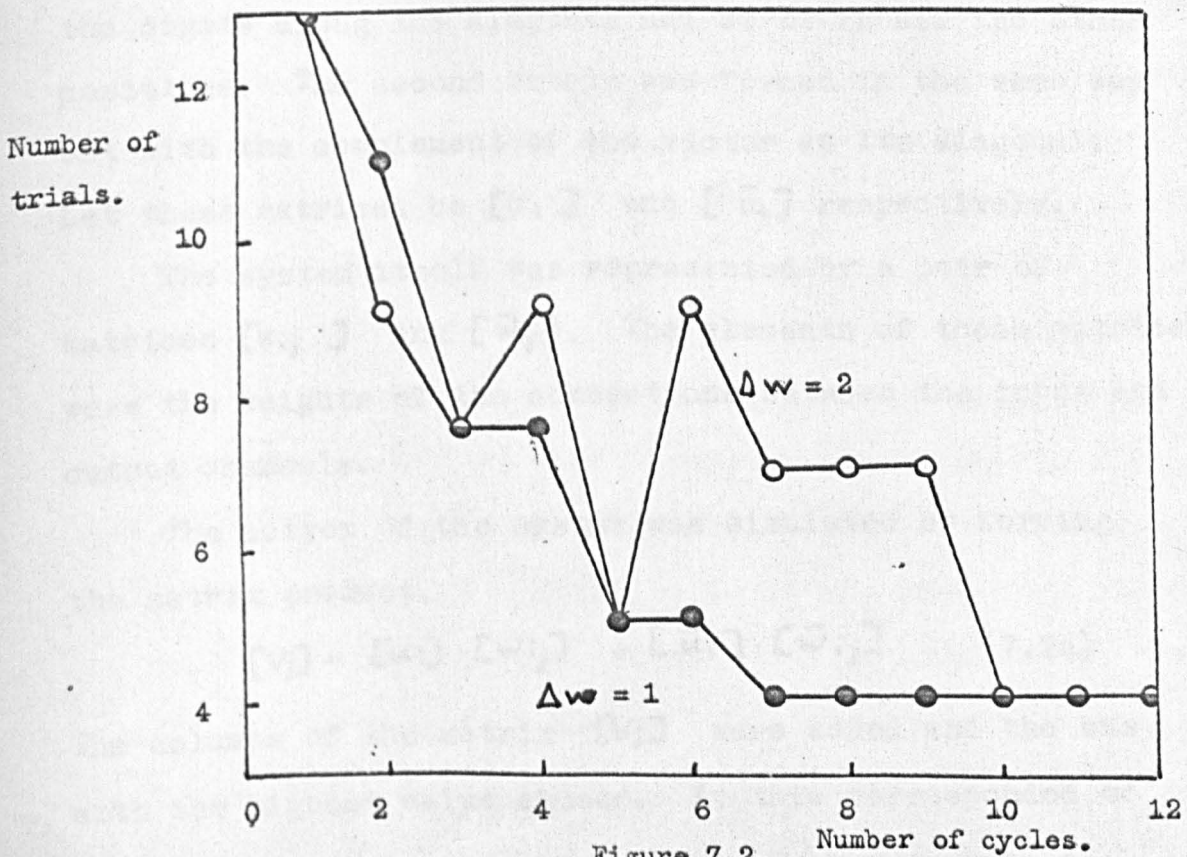


Figure 7.2.

This is convenient as c = 1 in equation (7.10) gives zero change in the mean weight of an ideal system. This may be interpreted as indicating that the mean of all the weights will remain unchanged so that full use will be made of their range. It should, therefore, be unnecessary to introduce any normalising procedure to ensure this.

7(4) Simulation of the system.

In order to test the adaptive procedure outlined above a simulation of the system was carried out. The input signals were represented by a pair of matrices. The first of these was formed by writing the pattern of input signals as a vector containing n binary digits. This vector was transformed into a matrix by writing the digits along the diagonal and zeros in all the other positions. The second matrix was formed in the same way but with the complement of the vector as its diagonal. Let these matrices be $[u_{i}]$ and $[\bar{u}_{i}]$ respectively.

The system itself was represented by a pair of matrices $[w_{ij}]$ and $[\bar{w}_{ij}]$. The elements of these matrices were the weights of the connections between the input and output channels.

The action of the system was simulated by forming the matrix product,

$$[v_j] = [u_i] \cdot [w_{ij}] + [\bar{u}_i] \cdot [\bar{w}_{ij}] \qquad (7.24)$$

The columns of the matrix $[v_j]$ were added and the one with the highest value chosen. If this corresponded to

the desired output, the appropriate weights were adjusted according to equation (7.6a), and if not equation (7.6b).

4 by 4 matrices were used and c was set equal to unity. The weights were determined initially by choosing at random integers between 1 and 10. The value of w should have been in the range,

$$\frac{9}{8} \leq \Delta w \leq \frac{9}{6} .$$

To ease the computations Δ w was made an integer. Two runs were tried with $\Delta w = 1$ and $\Delta w = 2$.

Four 4-digit patterns were presented. It was found that if the patterns were presented regularly it was possible for the values of some of the weights to become cyclic and this prevented the system from converging a state in which it could make the desired discriminations. Another mode of presentation was therefore adopted. A single pattern was presented repeatedly until the system gave the desired response. Then the next pattern was likewise presented.

With this method the system eventually converged to the required behaviour. Its performance is shown in Figure (7.2). The number of trials per cycle is plotted against the number of cycles. A cycle was completed when all four patterns had been recognised correctly once.

It can be seen from Figure (7.2) that the system took less trials to reach the desired behaviour with

$\Delta w = 1$ than with $\Delta w = 2$. This was as expected from the foregoing theory as $\Delta w = 1$ is nearer to the predicted optimum value of Δw. With $\Delta w = 2$ the 'error' curve did not continuously decrease, but showed wide fluctuations during part of the training period.

7 (5) Electrical circuit design.

An electrical circuit to perform these computations automatically has been constructed. The circuit had the form of a learning matrix (Figure 7.3(a)). The input signals were applied to the machine by means of the bipolar switches $S_1 \ldots S_n$. The rows of the resistance matrix were connected through resistors $r_1 \ldots r_m$ to earth. These rows were also connected to a device which indicated the row with the maximum voltage.

The circuit of Figure 7.3(b) is approximately equivalent to that of Figure 7.3(a). The approximation is equivalent to assuming that the columns which are not connected directly to the applied voltage V are all at some other potential V_o . As will be shown later the conditions are chosen so that this is very nearly true.

The resistances in Figure 7.3(b) are related to those in Figure 7.3(a) by,

$$\frac{1}{r_{ij}} = \sum_j [u_i]\left[\frac{1}{r_{ij}}\right] + \sum_j [\bar{u}_i]\left[\frac{1}{r_{ij}}\right] ,$$

$$\frac{1}{r_{oj}} = \sum_i [u_i]\left[\frac{1}{r_{ij}}\right] + \sum_j [\bar{u}_i]\left[\frac{1}{r_{ij}}\right] . \qquad (7.25)$$

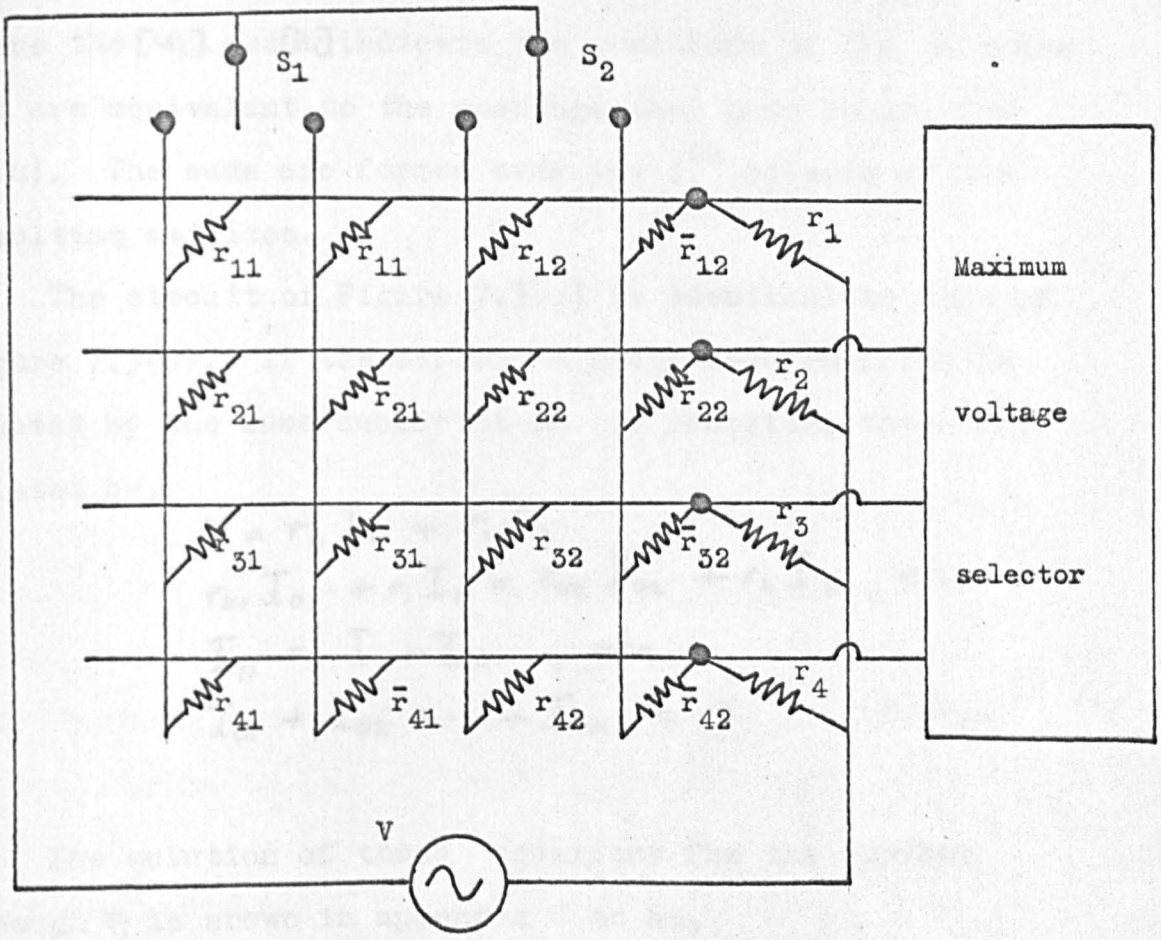


Figure 7.3(a)

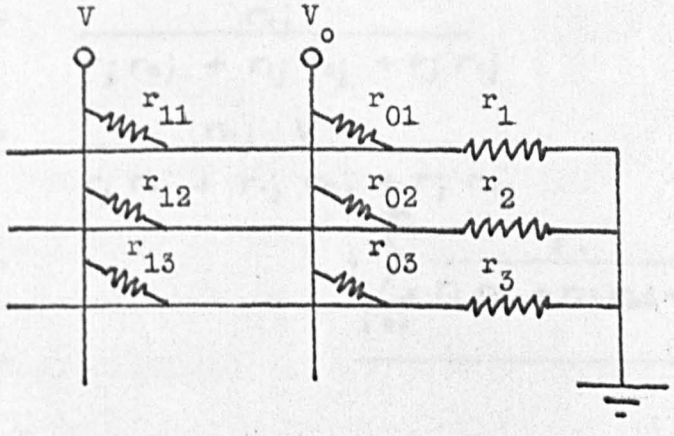


Figure 7.3(b).

where the [u;] and [ū;] indicate the positions of the switches and are equivalent to the meanings they have in section (7.4). The sums are formed over the j^{th} columns of the resulting matrices.

The circuit of Figure 7.3(c) is identical to that of Figure 7.3(b). If the current through each resistor is denoted by the same subscript as the resistor, these are related by,

$$V = r_{11} I_{11} + r_1 I_1$$
$$r_{01} I_{01} + r_1 I_1 = r_{02} I_{02} + r_2 I_2 \text{ , etc.}$$
$$I_{11} = I_1 - I_{01} \text{ , etc.}$$
$$I_{01} + I_{02} + \cdots + I_{0M} = 0. \qquad (7.26).$$

The solution of these equations for the current through r_j is shown in appendix C to be,

$$I_j = A_j V_0 + B_j \qquad (7.27)$$

where
$$A_j = \frac{r_{ij}}{r_j r_{oj} + r_{ij} r_{oj} + r_j r_{ij}}$$

$$B_j = \frac{r_{oj} V}{r_j r_{oj} + r_{ij} r_{oj} + r_j r_{ij}}$$

$$\frac{V_0}{V} = \frac{\sum_{i=1}^{m} \frac{r_i}{r_i r_{oi} + r_{ii} r_{oi} + r_i r_{ii}}}{\sum_{i=1}^{m} \frac{r_i + r_{ii}}{r_i r_{oi} + r_{ii} r_{oi} + r_i r_{ii}}}$$

For the system to perform the required computation, put $r_1 = r_2 = \ldots r_m = r$, and $r \ll r_{ij}, r_{oj}$. (7.28)

Equation (7.27) then reduces to,

$$I = \frac{V_o}{r_{oj}} + \frac{V}{r_{ij}} \qquad (7.29)$$

As $r \ll r_{ij}, r_{oj}$, $V_o \ll V$, so equation (7.29) may be written $I_j = \frac{V}{r_{ij}}$ (7.30).

Also as $V_o \ll V$, the potential of each column not connected directly to V will be approximately zero, and hence they will all be approximately equal.

The voltage across r_j will be,

$$V_j = rI_j = \frac{rV}{r_{ij}} \qquad (7.31)$$

Substututing equation (7.25) in equation (7.31) and writing $\frac{1}{r_{ij}} = W_{ij}$. W_{ij} is the conductance of the junction.

$$V_j = V r \left(\sum_j [u_i][w_{ij}] + \sum_j [\bar{u}_i][\bar{w}_{ij}]\right) \qquad (7.32)$$

This equation is very similar to equation (7.24).

This electrical circuit thus provides a method of realising physically the system of section (7.4). The conductances of the resistors in the matrix correspond to the weightsof that section.

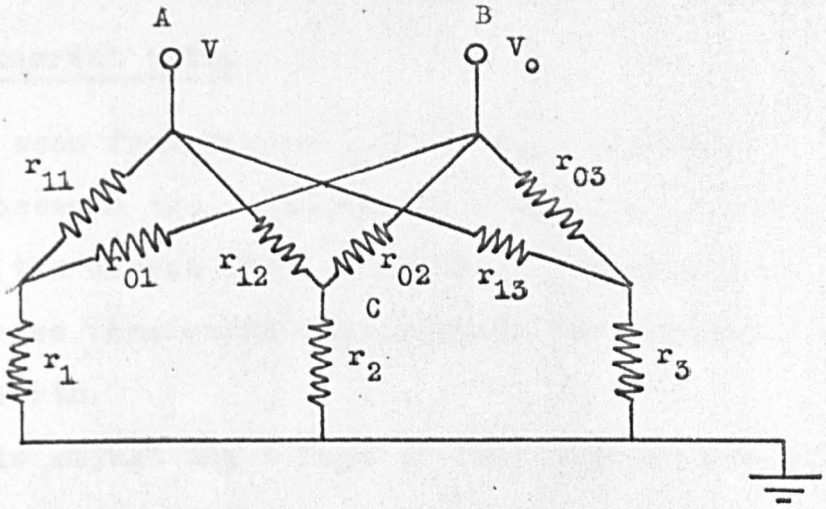


Figure 7.3(c).

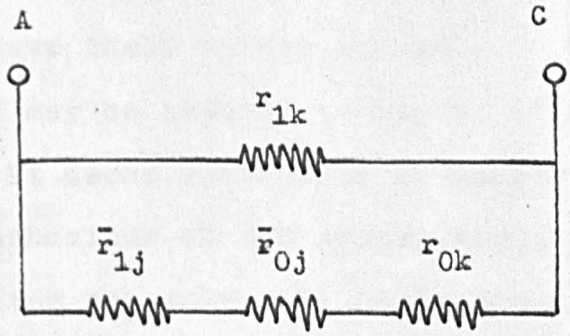


Figure 7.4.

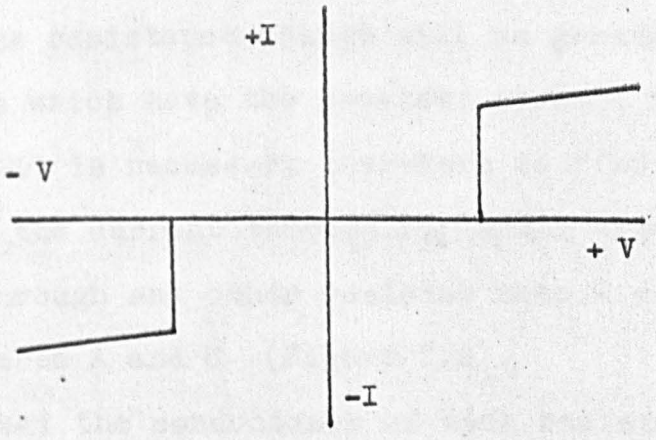


Figure 7.5.

7 (6) 'Stray' current paths.

As can be seen from Figure 7.3(a) there are many current paths between the input column i and the output row j besides the direct one through r_{ij}. Condition (7.28) makes these irrelevant when signals are applied between V and earth.

In order to adjust the values of the conductances of the junctions it is necessary to apply D.C. voltages between single rows and columns directly. In this case the stray currents are important as then other resistors besides r_{ij} will have their values changed.

Figure 7.3 (a) may be reduced to Figure (7.4) in this instance. It seems reasonable to suppose that the change in the behaviour of the system brought about by attempting to alter r_{ij} only will be in the direction desired only if the resistance r_{ij} is changed more than each of the others.

If electrolytic variable resistors are employed in this circuit the resistance change will be greatest in those resistors which have the greatest current flowing through them. It is necessary therefore to find the condition that the current through r_{ij} shall always be greater than through any other resistor when a potential is applied between A and C (Figure 7.4).

Suppose that the conductance of each resistor may

only vary between w_1 (high conductance) and w_2 (low conductance). For the limiting condition suppose that all the resistors which comprise r_{1k} have conductance w_2 and all the others w_1.

Hence
$$\left.\begin{aligned} r_{1k} &= \frac{1}{n w_2} \\ r_{ok} &= \frac{1}{n w_1} \\ \bar{r}_{ij} &= \frac{1}{(m-1)\, n w_1} \\ \bar{r}_{oj} &= \frac{1}{(m-1)\, n w_1} \end{aligned}\right\} \quad (7.34)$$

The required condition is,

$$r_{1k} < \bar{r}_{ij} + \bar{r}_{oj} + r_{ok}$$

$$\frac{1}{m w_2} < \frac{1}{m w_1} + \frac{2}{n(m-1) w_1}$$

or
$$w_2 > \left(\frac{m-1}{m+1}\right) w_1 \quad (7.35)$$

Hence w_{ij} must always be within the range,

$$\left(\frac{m-1}{m+1}\right) w_1 \leq w_{ij} \leq w_1 \quad (7.36)$$

where w_1 is the maximum conductance it may assume.

7(7) Discriminating power.

These limits placed on the conductances of the variable resistors may be used to calculate the resolving

power that it is necessary for the maximum voltage selector to possess. Alternatively, if this is fixed and known, they may be used to calculate the discriminating power of the matrix, or the maximum size of matrix of which full use could be made.

Suppose that the resistors are adjusted so that the k th row is at the highest possible voltage for a given input pattern, then,

$$r_{1k} = \frac{1}{nw_1}$$

and so $$V_k = rV\, w_1 n.$$

Suppose another input pattern differing by a Hamming distance d is applied,

$$V_k' = rV\{(n-d)\, w_1 + d w_2\}$$

for a fully trained system. The voltage selector needs to be able to discriminate between these voltages, therefore the resolving power,

$$\frac{\Delta V}{V} \geqslant \frac{V_k - V_k'}{V} \geqslant r(w_1 - w_2)\, d. \qquad (7.37)$$

This is the resolving power required for a voltage selector coupled to a fully trained system. For a system in which adaptation is taking place the voltage selector needs to be about $\frac{w_1 - w_2}{\Delta w}$ more times sensitive than this.

The discriminating power of a matrix may be expressed

as $\frac{1}{d_m}$, where d_m is the Hamming distance between the most similar patterns that it can distinguish.

$$\frac{1}{d_m} = \frac{r(w_1 - w_2)}{(R.P.)} \qquad (7.38)$$

The maximum discriminating power occurs when $d = 1$, $\therefore$ $R.P. = r(w_1 - w_2)$.

But, from equation (7.36) $w_2 = \frac{m-1}{(m+1)}$

Hence $$m = \frac{rw_1}{(R.P.)} - 1 \qquad (7.39)$$

m is the largest number of output channels which a matrix of this type should possess.

7 (8) Construction of machine.

A machine was built to the design given in the earlier parts of this chapter. The adaptive components were four nine-terminal devices similar to that illustrated in Figure (3.5). The holes had a bore of approximately 0.5 mm. These devices were filled with 1.0 gm. per c.c. silver nitrate solution, and sealed with 'Araldite' epoxy resin. Some difficulty was encountered in sealing the tubes. The silver nitrate solution reacted with the sealing compound if it came into contact with it before it had hardened, so the following procedure had to be adopted. The tubes and the bottom plate were first glued on to the perspex body.

Several days later this was filled with electrolyte. The tubes were set at such an angle that the surface tension forces caused capillary rise in the tubes so that the electrolyte almost, but not quite, reached the ends. It was then possible to seal a silver electrode in each tube and to fix the lid containing the common electrode on top without the electrolyte and the sealing compound making contact.

The variable resistance devices were stacked vertically, and the limiting resistances (which were calculated from equation (7.36)) were attached.

The mean resistances of the tubes was about 10 kilohms, so in order to satisfy condition (7.28) the resistors $r_1 \ldots r_4$ (Figure (7.3) were given the value 100 ohms.

The variable resistance matrix was finally connected into the circuit of Figure (7.6). The pattern of signals presented to the system was determined by the positions of the switches on the input channels of the resistance matrix. With the switch S in one position A.C. signals passed through the matrix and the output channel which was at the highest potential was indicated. With the switch S in its other position D.C. signals could be made to pass between selected junctions by the trainer.

7 (9) Preliminary investigations.

An attempt was made to train the system to recognise

four patterns after the manner of section (7.4). After a large number of trials (about 100) the performance of the machine had not improved. Measurements of the conductances of the tubes after successive trials showed that there were no signs of these converging to the expected values. It was concluded that the 'stray' currents so interfered with the conductances that they changed their values more than the currents intended to change them did.

Although the limiting resistors (section 7.8 and equation 7.36) ensured that this did not occur during a given trial, each tube had its conductance changed deliberately, on average, only one out of every eight trials. The total change produced by the 'stray' currents during the seven trials was greater than the deliberate change produced on the eighth.

This being the case, attempts to train the present machine by this trial-and-error method were discontinued. Instead the machine was trained by a method similar to that used by Steinbuch with his learning matrices (v. Chapter VI). All conductances were initially made low, then some of them were increased by passing currents between the inputs and the desired output. In this way the machine was adjusted to distinguish between four different four-bit patterns.

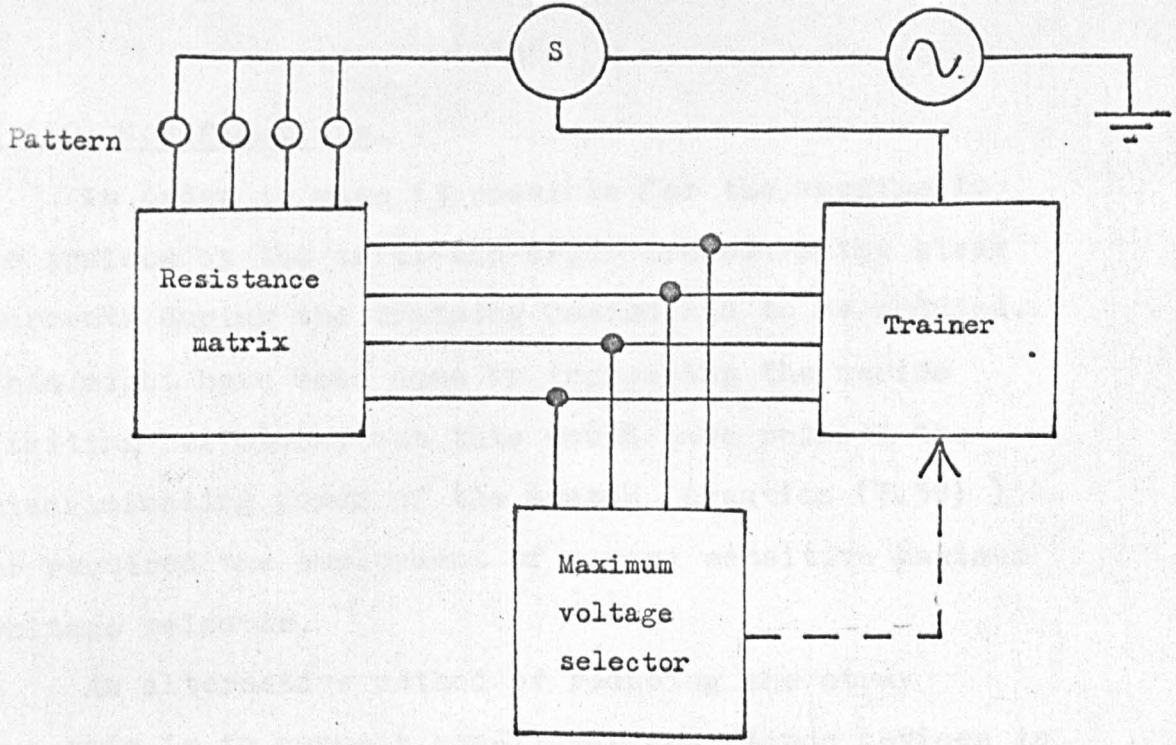


Figure 7.6.

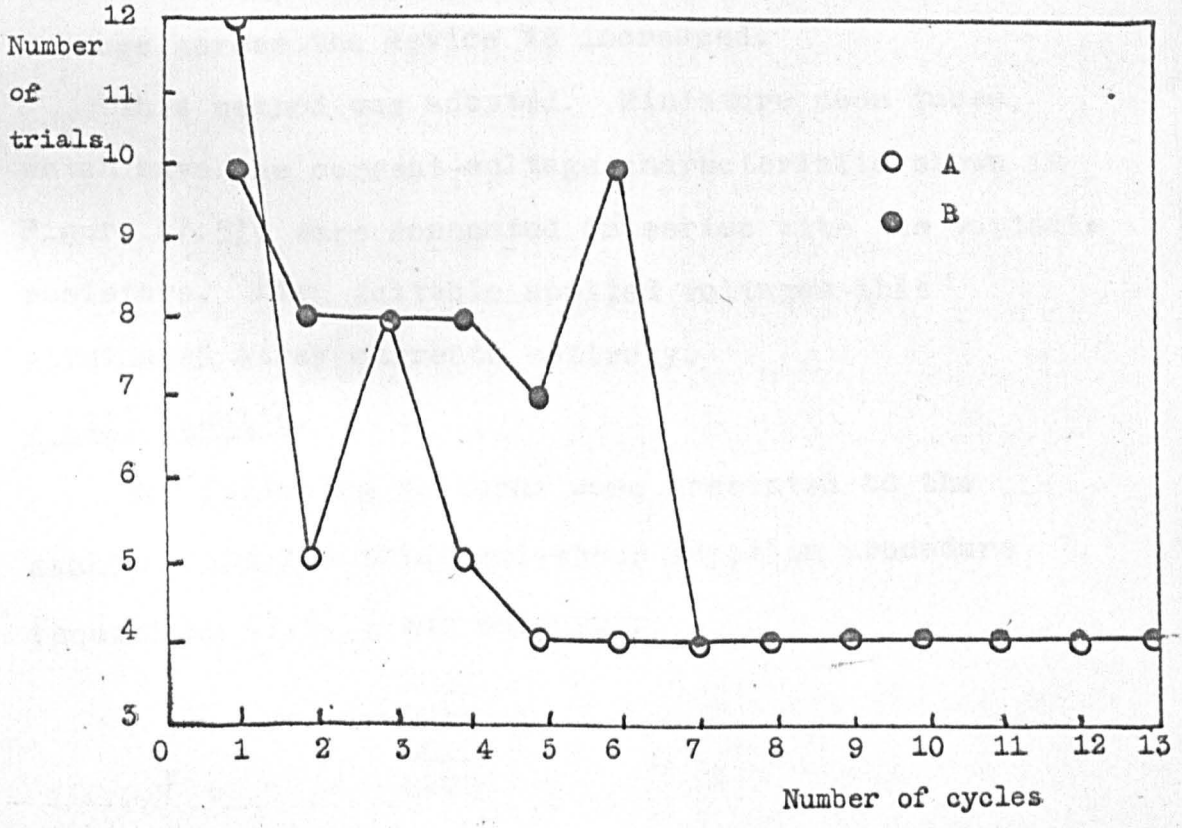


Figure 7.7.

7 (10) Modifications.

In order to make it possible for the machine to be trained by the trial-and-error procedure the stray currents during the training period had to be reduced. This might have been done by increasing the series limiting resistors, but this would have reduced the discriminating power of the system (equation (7.38)) or required the employment of a very sensitive maximum voltage selector.

An alternative method of reducing the stray currents is to connect non-linear resistance devices in series with the adaptive components in place of the limiting resistors. The type of non-linearity required is one in which the resistance becomes less as the voltage across the device is increased.

This method was adopted. Miniature neon tubes, which have the current-voltage characteristic shown in Figure (7.5), were connected in series with the variable resistors. With suitable applied voltages this eliminated stray currents entirely.

7 (11) Results.

The following patterns were presented to the machine, and the trial-and-error adoption procedure (equations (7.6)) was employed:

0001
0010
0011
0100

After some 34 trials the system could distinguish between these various patterns. The 'learning curve' (A of Figure 7.7) shows the speed with which this was brought about. The number of trials per cycle was plotted against the number of cycles.

Some time later the system was adapted to recognise the following patterns instead

1 0 0 0
1 1 0 0
1 1 1 0
1 1 1 1

It took about 50 trials for the system to respond correctly every time to this set (B of Figure 7.5). This was considerably longer than with the first set, but it is expected that there should be large differences in the number of trials required for adaption as this depends very much on the initial conductances.

It was notice that the machine actually took fewer trials to adapt than were required in each of the paper simulations. (Figure 7.2). The most perfect machine which stored all the data and had facilities for correlating it would take, on average, 16 trials.

7 (12) Discussion.

The adaptive system described in this chapter was constructed to show that the devices of Chapters III and V

could be used in the fabrication of adaptive systems. The anodic oxide film cells could probably be used instead of the dendrite cells but for reasons given in Chapter V design of the circuit would be more difficult.

The system employing the dendrite cells worked surprisingly well in view of the fact that it contained several sub-standard components. One of the tubes was later found to be open circuit all the time, and there was 100% variation in the maximum resistances of the tubes, and 20% variation in the running voltages of the neons. Also air bubbles developed in some of the tubes.

The parameters Δw and c were chosen as intelligently as possible from equation (7.23). An approximately constant current (4 - 6 milliamps) was applied for a constant time in order to adjust each resistance, but as there was some variation in the growth of a dendrite over a short period of time this did not always lead to the same change in resistance.

These facts illustrate a principle that is gradually emerging; that adaptive systems can, provided they are given occasional training periods, be made to function as well as perfect fixed systems even though they suffer from certain component failures.

CHAPTER VIII

CONCLUSIONS AND FURTHER POSSIBILITIES

8 (1). The growth of silver dendrites.

The most important practical result to emerge from the study of silver dendrites is that when restricted to narrow tubes they can be used reliably to control a variable resistance, enabling it to be increased as well as decreased.

Among further developments suggested by these findings it would seem that growth in even smaller bore tubes should be investigated. As well as reducing the physical size and power consumption of the devices, this would increase the maximum rate of growth of the dendrites and so increase their field of potential application. The generalisation of this device to produce non-linear relations between resistance and time may also prove to be of consequence. *

8 (2) Anodic oxide films.

The investigation of the properties of anodic oxide films has shown that it is possible to control their

*Another extension of this work is possible. Under certain conditions the ionic conductivity of silver iodide crystals is abnormally high (Mott and Gurney,1948). This suggests that it might be possible to grow and dissolve silver dendrites inside these crystals, in which case miniature solid-state variable resistors could be constructed. Such devices would probably be free from the disadvantages caused by polarisation which are inevitably present when liquid electrolytes are employed.

impedance in a reproducible manner by the quantity of electric charge passed through them. In Chapter IV it has been shown that if the impedance is altered by electrochemical dissolution of the film, under constant voltage or current, an ever diminishing rate of change of impedance with time is to be expected instead of an accelerating one as might at first have been supposed.

If the size of devices utilising this process is reduced care must be taken to avoid evaporating the small bulk of electrolyte by the relatively large currents passed during erosion of the film. If, however, these devices are only used for uni-directional impedance changes, miniaturisation will be easier to implement, and may prove particularly useful if the parallel switching principle (Chapter V) is to be employed, as then there will be only small quantities of liquid to be moved. With small scale devices physico-chemical phenomena, such as evaporation, surface tension and electro-osmosis, might be employed to alter the distribution of the electrolyte.

8 (3) Adaptive systems.

The system described in Chapter VII operated well enough to show that the electrolytic variable resistors which had been developed are suitable for use in an adaptive machine.

The training procedure developed for this system is also interesting in its own right. The behaviour of the

system depends upon the values of the parameters Δw and c . By analogy with other adaptive systems it seems likely that gradual variation of these parameters on the basis of performance during the training period would cause the system to adapt to a new set of patterns more rapidly. In future work the effect of some of these changes on the performance of the system should be examined.

Another training routine which suggests itself for future investigation is one in which the patterns are presented in random order. For machines in which adaptive matrices are used in self-organising systems to cope with unpredictable changes (v. section 8.5) this would be more appropriate than the procedure outlined in Chapter VII.

A further line of investigation which looks fruitful would be to consider the properties of two or more adaptive matrices connected in series. There are, however, a number of problems, mentioned in Chapter VI, in extending the training procedure to cope with this situation.

Parallel systems in which two or more adaptive matrices are connected between the same set of input and output channels should also be interesting, since if oxide films are employed as the adaptive components, full benefit of the switching principle (Chapter V) could be obtained. The matrix active at a given time could be selected by another matrix which had been taught to abstract the

'context' of the pattern. In this way a hierarchical structure of adaptive organisers similar to that proposed by MacKay (1956) could be built up.

8 (4) Self-organising systems.

The training procedure of Chapter VII was adopted partly because it offered the possibility of using the adaptive matrix in a self-organising system. The adaptive system plus the trainer does in fact show all the features of self-organisation mentioned in Chapter I. The goal of the system is to produce the output desired by the trainer for a given input pattern; the evaluative mechanism is the trainer who indicates the correctness or otherwise of the response; and the source of variation is the automatic adjustment of the weights as a result of the trainer's signals.

By replacing the trainer with some device which measures the 'distance' of the system from its goal a fully automatic self-organising system can be produced. The input to the adaptive matrix could be a pattern of signals representing the state of the world in which the system is situated. This input pattern would produce an output leading to an action. If this action brought the system nearer to its goal the weights could be increased automatically. In this way the system could learn for itself how to attain its goal from any position it had been in before.

One of the difficulties with the adaptive system and training procedure of Chapter VII is that even if the training procedure could be adjusted so that the machine behaved ideally, the training period would still increase rapidly as the number of output channels (actions) is increased. One of the ways of reducing this period is to employ the 'maturation' principle (MacKay, 1956, 1961). MacKay suggested that the repertoire of a self-organising system should initially be small, and that as the system learned to perform its actions correctly this repertoire should be progressively increased.

8 (5) Applications.

Although self-organising systems appear to be capable of performing remarkable feats, there are only a limited number of situations in which it is economical to employ them in practice. If the task to be performed is exactly specifiable it is best to use a 'fixed' machine. If, however, the environment in which the system is working is subject to indeterminate changes a fixed machine may find itself in a situation in which it can no longer cope. An adaptive system is therefore essential.

If the transfer function of the machine itself changes, owing to deterioration of some of its components, a non-adaptive system may become incapable of performing its task. A self-organising system can adjust its own transfer function on the basis of its performance so that

its lifetime can be made much longer than that of a fixed system even though it is built of components of the same reliability.

Besides helping to further the construction of artificial self-organising systems there are a number of more specific uses to which electrolytic growth processes might be put. For example, some of the adaptive control systems for industrial processes mentioned by Aseltine et al (1958) could be constructed using these devices provided that the time constants of the processes to be controlled were not too short,.

The field of teaching machines is another area in which the electrolytic devices might be used. Present teaching machines seem to be small and suited only to teaching simple tasks (Lumsdaine and Glaser, 1960) or so complex that a whole digital computer is required for each student (Smallwood, 1962). This latter type use the computer for calculating and storing conditional probabilities. The devices of Chapter III could be employed to do this with advantage.

Other possible applications of electrolytic growth processes include making use of the growth of silver dendrites in automatic wiring of machines, and employing the devices of Chapter III, in which the length of a dendrite is proportional to the time integral of the charge passed, as inexpensive analogue substitutes for

electromechanical timers and counters in situations in which high accuracy is not required.

8 (6) Conclusion.

This thesis has concentrated on those properties of electrolytic growth processes which relate to their usefulness in the construction of complex, adjustable automata. The results obtained show that the growth of conductors and the growth of insulation offer satisfactory and inexpensive means of determining large numbers of variable parameters in adaptive systems.

REFERENCES, CHAPTER VIII

Aseltine, J.A., Macini, A.R., and Sature, C.W. (1958) "A Survey of Adaptive Control Systems." I.R.E. Trans. on Automatic Control, Dec., 1958., pp 102-8.

Lumsdaine, A.A. and Glaser, R. (Eds) (1960) "Teaching Machines and Programmed Learning." National Education Association, United States.

MacKay, D.M. (1956) "The Epistemological Problem for Automata" in Automata Studies, Princeton University Press.

MacKay, D.M. (1961) "Information and Learning" in "Lernende Automaten", Oldenburg, Munich.

Mott, N.F. and Gurney, R.W. (1948) "Electronic Processes in Ionic Crystals" Oxford, p. 62.

Smallwood, R.D. (1962) "A Decision Structure for Teaching Machines." M.I.T. Press, Cambridge, Mass.

APPENDIX A

With reference to Figure A.1.,

r_1'' and r_2'' are the internal and external radii of the image of the tube on the screen,

r_1 and r_2 are the true internal and external radii of the tube

r_1' is the apparent internal radius.

$$r_1' = r_2 \sin\beta \qquad (A.1)$$

Now $$\frac{r_1}{\sin\alpha} = \frac{r_2}{\sin(\pi/2 - \alpha + \beta)} \qquad (A.2)$$

$$\therefore \frac{r_2}{r_1} \sin\alpha = \cos\alpha\cos\beta + \sin\alpha\sin\beta . \qquad (A.3)$$

But, $$\frac{\sin\beta}{\sin\alpha} = \mu \quad \text{(refractive index)} \qquad (A.4)$$

$$\therefore \cos\alpha = \left\{ 1 - \frac{\sin^2\beta}{\mu^2} \right\}^{1/2}$$

Substitute in (A.3)

$$\frac{r_2}{r_1}\frac{\sin\beta}{\mu} = \left\{ 1 - \frac{\sin^2\beta}{\mu^2} \right\}^{1/2} \cos\beta + \frac{\sin^2\beta}{\mu} .$$

$$\therefore r_1 = \frac{r_2 \sin\beta}{(\mu^2 - \sin^2\beta)^{1/2}\cos\beta + \sin^2\beta} \qquad (A.5)$$

From (A.1) and (A.5)

$$r_1 = \frac{r_1'}{(\mu^2 - \sin^2\beta)^{1/2}\cos\beta + \sin^2\beta}$$

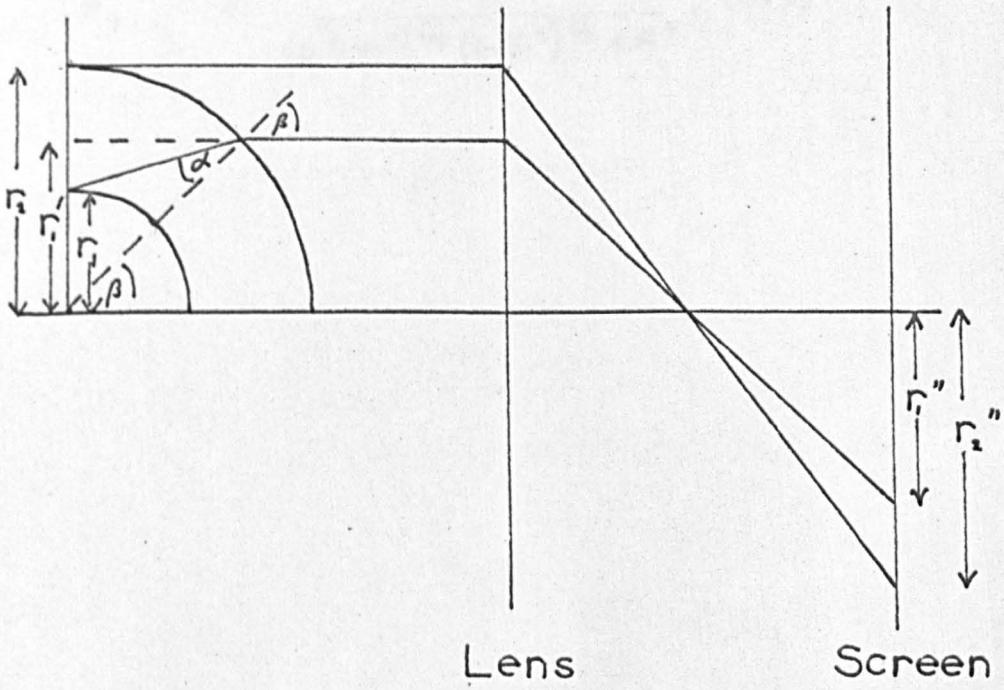


Figure A.1.

But, $\sin\beta = \frac{r_1'}{r_2} = \frac{r_1''}{r_2''} = \bar{m}$

Hence $r_1 = \frac{r_1'}{(\mu^2 - \bar{m}^2)^{1/2}(1-\bar{m}^2)^{1/2} + \bar{m}^2}$. (A.6)

APPENDIX B

The conditional probability that the signal in the j^{th} output channel is v_j when the input is (u_i) is given by Bayes' rule as (equation 6.2)

$$P(v_j / (u_i)) = \frac{P(v_j) \cdot P((u_i)/V_j)}{P((u_i))} \quad (B.1.)$$

Write $(u_i) = U_i$,

then $$P(v_j / U_i) = \frac{P(v_j) \cdot P(U_i / v_j)}{P(U_i)} \quad (B.2)$$

The signal in the j^{th} channel is either v_j or $\bar{v}_j$

$$\therefore \quad P(U_i) = P(v_j) \cdot P(U_i / v_j) + P(\bar{v}_j) \cdot P(U_i / \bar{v}_j)$$

$$\therefore \quad P(v_j / U_i) = \frac{P(v_j) \cdot P(U_i / v_j)}{P(v_j) \cdot P(U_i / v_j) + P(\bar{v}_j) P(U_i / \bar{v}_j)}$$

$$= \frac{1}{1 + \frac{P(\bar{v}_j)}{P(v_j)} \cdot \frac{P(U_i / \bar{v}_j)}{P(U_i / v_j)}} \quad (B.3)$$

If the signals in the input channels occur independently of each other,

$$P(U_i / v_j) = \prod_{i=1}^{n} P(u_i / v_j),$$

and $$P(U_i / \bar{v}_j) = \prod_{i=1}^{n} P(u_i / \bar{v}_j).$$

Hence $$P(v_j / U_i) = \frac{1}{1 + \frac{P(\bar{v}_j)}{P(v_j)} \prod_{i=1}^{n} \frac{P(u_i / \bar{v}_j)}{P(u_i / v_j)}} \quad (B.4)$$

But $P(u_i/v_j) = \dfrac{P(u_i) \cdot P(v_j/u_i)}{P(v_j)}$,

and $P(u_i/\bar{v}_j) = \dfrac{P(u_i) \cdot P(\bar{v}_j/u_i)}{P(\bar{v}_j)}$, (B.5)

by Baye's rule.

Hence, $$P(v_j/(u_i)) = \frac{1}{1 + \left[\frac{P(v_j)}{P(\bar{v}_j)}\right]^{n-1} \prod_{i=1}^{n} \frac{P(\bar{v}_j/u_i)}{P(v_j/u_i)}}. \quad (B.6)$$

which is equation (6.3).

APPENDIX C

The currents through the resistors in the circuit of Figure 7.3(c) are given by equation (7.26).

$$V = r_{11} I_{11} + r_1 I_1 \qquad (C.1)$$

$$r_{01} I_{01} + r_1 I_1 = r_{02} I_{02} + r_2 I_2 \text{ , etc.} \qquad (C.2)$$

$$I_{11} = I_1 - I_{01} \text{ , etc.} \qquad (C.3)$$

$$I_{01} + I_{02} + \cdots + I_{0M} = 0. \qquad (C.4)$$

From equations (C.1) and (C.3)

$$V = r_{11} (I_1 - I_{01}) + r_1 I_1$$

$$= (r_{11} + r_1) I_1 - r_{11} I_{01}$$

$$I_1 = \frac{V + r_{11} I_{01}}{r_1 + r_{11}} \qquad (C.5)$$

From equation (C.2),

$$r_{01} I_{01} + r_1 I_1 = r_{02} I_{02} + r_2 I_2 \text{ , etc.}$$

$$= V_0 \text{ (say).}$$

$$I_{01} = \frac{V_0 - r_1 I_1}{r_{01}} \qquad (C.6)$$

Substitute (C.6) in (C.5)

$$(r_1 + r_{11}) I_1 = V + \frac{r_{11} (V_0 - r_1 I_1)}{r_{01}}$$

$$I_1 = \frac{r_{01} V + r_{11} V_0}{r_1 r_{11} + r_{01} r_1 + r_{01} r_{11}} \qquad (C.7)$$

or $$I_1 = A_1 V_0 + B_1$$

where $$A_1 = \frac{r_{11}}{r_1 r_{11} + r_{01} r_1 + r_{01} r_{11}}$$

$$B_1 = \frac{r_{01} V}{r_1 r_{11} + r_{01} r_1 + r_{01} r_{11}}$$

Which is equation (7.27).

Substitute (C.5) in (C.6)

$$r_{01} I_{01} = V_0 - \frac{r_1 (V + r_{11} I_{01})}{r_1 + r_{11}}$$

$$I_{01}(r_1 r_{01} + r_{11} r_{01} + r_1 r_{11}) = (r_1 + r_{11}) V_0 - r_1 V$$

$$I_{01} = \frac{(r_1 + r_{11}) V_0 - r_1 V}{r_1 r_{01} + r_{11} r_{01} + r_1 r_{11}} \qquad (C.8)$$

Similar equations can be obtained for $I_{02} \ldots I_{0M}$.

Substitute these in equation (C.4)

$$V_0 \sum_{i=1}^{M} \frac{r_i + r_{ii}}{R_i} - V \sum_{i=1}^{M} \frac{r_i}{R_i} = 0,$$

where $$R_i = r_i r_{0i} + r_{ii} r_{0i} + r_i r_{ii}$$

Hence, $$\frac{V_0}{V} = \frac{\sum_{i=1}^{M} \frac{r_i}{R_i}}{\sum_{i=1}^{M} \frac{r_i + r_{ii}}{R_i}} \qquad (C.9).$$